Gene expression profiling of human lymph node-positive gastric adenocarcinomas:

–Towards personalized prognosis and therapy–

Dissertation

zur Erlangung des akademischen Grades
doctor rerum naturalium
(Dr. rer. nat.)
im Fach Biologie

eingereicht an der
Mathematisch-Naturwissenschaftlichen Fakultät I
der Humboldt-Universität zu Berlin

von

Dipl.-Ernähr. Susann Förster
geb. am 07.06.1980 in Leipzig

Präsident der Humboldt-Universität zu Berlin:
Prof. Dr. Dr. h.c. Christoph Markschies

Dekan der Mathematisch-Naturwissenschaftlichen Fakultät I:
Prof. Dr. Lutz-Helmut Schön

Gutachter: 1. Prof. Dr. Hanspeter Herzel
2. Dr. habil. Wolfgang Kemmner
3. Prof. Dr. Reinhold Schäfer

Tag der mündlichen Prüfung: 23.09.2010

Zusammenfassung

Histologisch werden zwei Haupttypen des Magenkarzinoms unterschieden, diffuse und intestinale Adenokarzinome. Beide Typen haben verschiedene Ätiologien, Abfolgen der Karzinogenese und Progression und im Zusammenhang damit verschiedene molekulare Profile.

Um den molekularen Hintergrund und das Verhalten beider Typen tiefgehender zu analysieren, wurden in der hier vorgelegten Arbeit globale Genexpressionsprofile mittels Affymetrix Microarray-Technik erstellt. Der intestinale Typ konnte als stark proliferierender Tumor mit signifikanter Überexpression von zellzyklus- und mitoserelevanten Genen definiert werden, während der diffuse Typ als stark stromaabhängig mit signifikanter Überexpression von Genen der extrazellulären Matrix hervortrat. Thrombospondin 4 (*THBS4*) wurde dabei als das am stärksten differentiell exprimierte Gen identifiziert. Es wird in diffusen Tumoren eminent überexprimiert, während keine Expression in intestinalen Tumoren zu verzeichnen ist. Immunhistochemische Studien bestätigten diese starke Überexpression auf Proteinebene und zeigten, dass THBS4 eine übermäßig angereicherte extrazelluläre Komponente des Tumorstromas ist. Intrinsische Expression von THBS4 in „gesundem“ Magenepithel und -stroma konnte nicht festgestellt werden. Kolokalisierungsstudien zeigten zudem, dass THBS4-positive Zellen auch positiv für Vimentin und α-Smooth muscle actin sind. Diese Ergebnisse belegen, dass THBS4 von Subpopulationen Tumor-assoziierter Fibroblasten (TAF) exprimiert und sezerniert wird. Dies konnte durch zusätzliche *in vitro* Experimente bestätigt werden, die aufzeigten, dass TAF von diffusen Tumoren eine ausgeprägtere *THBS4*-mRNA Expression aufweisen als normale Fibroblasten des Magens gleicher Patienten. Abschließend konnte in *in vitro* Kokultur-Studien aufgedeckt werden, dass die *THBS4*-Expression in Fibroblasten durch Tumorzellen diffuser Magenadenokarzinome transkriptionell stimuliert und erhöht wird. Die vorliegende

Arbeit ist die erste, die THBS4 im Szenario diffuser Magenadenokarzinome beschreibt und charakterisiert.

Metastasenbefall regionaler Lymphknoten (N+) ist bei den meisten Magenadenokarzinomdiagnosen bereits vorhanden. Dieser ist der stärkste derzeit verfügbare Parameter zur Abschätzung der Prognose und Wahl der Therapie. Allerdings reicht er allein für eine eindeutige Prognosebestimmung nicht aus. Um ergänzende molekulare Prognoseindikatoren zu identifizieren, wurden aus den Microarray-Daten dieser Studie Gene, deren Expression mit dem klinischen Verlauf von N+ Patienten korreliert, extrahiert. Einige der ermittelten Gene, wie RAN binding protein 17, homeobox C10, ras-related associated with diabetes und folate receptor 1, konnten mittels quantitativer real-time PCR als prognoserelevant validiert werden. Eine signifikante Stratifizierung der N+ Patienten bezüglich progressionsfreiem Überleben war anhand der Expression dieser Gene möglich. Solche molekularen Marker oder Signaturen können in Zukunft dazu dienen, spezifischere personenbezogene Prognosevorhersagen zu treffen und auf diese Weise eine auf den jeweiligen Patienten optimierte Therapie auszuwählen.

Schlagworte: Genexpressionsprofilierung, Magenadenokarzinom, Tumorstroma, extrazelluläre Matrix, Tumor-assoziierte Fibroblasten, N-Stadium, klinischer Verlauf, Prognose

Abstract

According to histology, two major subtypes of gastric cancer can be distinguished, diffuse and intestinal-type adenocarcinomas. They are assumed to have distinct etiologies, follow different pathways of carcinogenesis and progression and, along with that, possess different molecular profiles.

To work towards a better understanding of each type's molecular background and biological behavior, global gene expression profiles were established in the work presented here, using the Affymetrix microarray technique. The intestinal type was identified to be a highly proliferative entity with significant overexpression of cell cycle and mitosis-relevant genes, whereas the diffuse type was proven to be strongly stroma-dependent with significant overexpression of extracellular matrix genes. Thrombospondin 4 (*THBS4*) was identified as the gene most differentially expressed between the two types. It is vastly overexpressed in diffuse-type tumors, whereas intestinal tumors lack its expression. Immunohistochemical studies proved overexpression on protein level and elucidated that THBS4 is a heavily accumulated extracellular constituent of the tumor stroma. Intrinsic THBS4 expression in "healthy" gastric epithelium and stroma was not encountered. Colocalization studies uncovered that THBS4-positive cells are also positive for vimentin and α-smooth muscle actin. These data signify that THBS4 is expressed and secreted by subpopulations of cancer-associated fibroblasts (CAFs). This was further evidenced by *in vitro* experiments demonstrating that *THBS4* mRNA expression is increased in CAFs of diffuse-type tumors compared to "matched" normal gastric fibroblasts. Finally, *in vitro* coculture studies revealed that transcriptional *THBS4* expression in fibroblasts is stimulated and increased by diffuse-type gastric tumor cells. This study is the first to describe and characterize THBS4 in the scenario of diffuse-type gastric adenocarcinomas.

Metastatic involvement of regional lymph nodes (N+) usually accompanies diagnosis of gastric adenocarcinoma and is currently considered the most important

parameter for assessment of prognosis and choice of therapy. However, estimation of prognosis based on this parameter alone is not sufficiently reliable. In order to identify additional molecular prognosis markers, genes whose expression correlates with clinical outcome of N+ patients were extracted from the microarray data set of this study. Via quantitative real-time PCR, several genes, including RAN binding protein 17, homeobox C10, ras-related associated with diabetes and folate receptor 1, were successfully validated to be prognosis-relevant. A significant gene expression-based stratification of the N+ patients with respect to disease-free survival was possible with these genes. Such molecular markers or signatures for prognosis may lead the way to more accurate and personalized prognosis assessment and, consequently, to more tailored therapeutic approaches in future.

Keywords: gene expression profiling, gastric adenocarcinoma, tumor stroma, extracellular matrix, cancer-associated fibroblasts, N-stage, clinical outcome, prognosis

Table of contents

List of figures

List of tables

1 Introduction

1.1 Gastric cancer incidence and mortality

Although the incidence of gastric cancer has been declining significantly around the world over the last decades (Munoz and Franceschi 1997, National Cancer Institute 2002, Parker *et al*. 1997, Wayman *et al*. 2001), it still represents the fourth most common malignancy worldwide and accounts for ~10% of all cancer-related deaths (American Cancer Society 2007, IARC 2008, Parkin *et al*. 2005). In general, stomach cancer rates are about twice as high in males as in females. Thus, it ranks at position two among the most common cancers in males and at position four in females (Fig. 1).

Gastric cancer incidence varies greatly across geographic locations with largest numbers found in Japan, Korea, China and certain countries of Eastern Europe and Latin America. In Western Europe, the highest incidence rate can be encountered in Portugal (Fig. 2).

Despite declining incidence, the survival from gastric cancer remains unchanged and poor, making it the second most common cause of cancer-related death worldwide (Parkin *et al*. 2005). The main reason for this poor survival is the rather late detection of the disease at advanced stages in most countries (Dicken *et al*. 2005, Hundahl *et al*. 2000, Siewert *et al*. 1998). One exception is Japan, where mass screening for gastric cancer and its pre-cancerous lesions is practiced and therefore most cancers are detected at early stages. This early detection screening routine, which is practiced since the 1960s, has enabled Japan to increase survival rates up to 52% (American Cancer Society 2007, Lee *et al*. 2006). Similar screening systems have recently emerged in Korea, Venezuela, Chile and Costa Rica. Nevertheless, survival from gastric cancer is poor in the vast majority of countries

with five-year relative survival rates of about 20–25% in Europe and the United States for instance (American Cancer Society 2007).

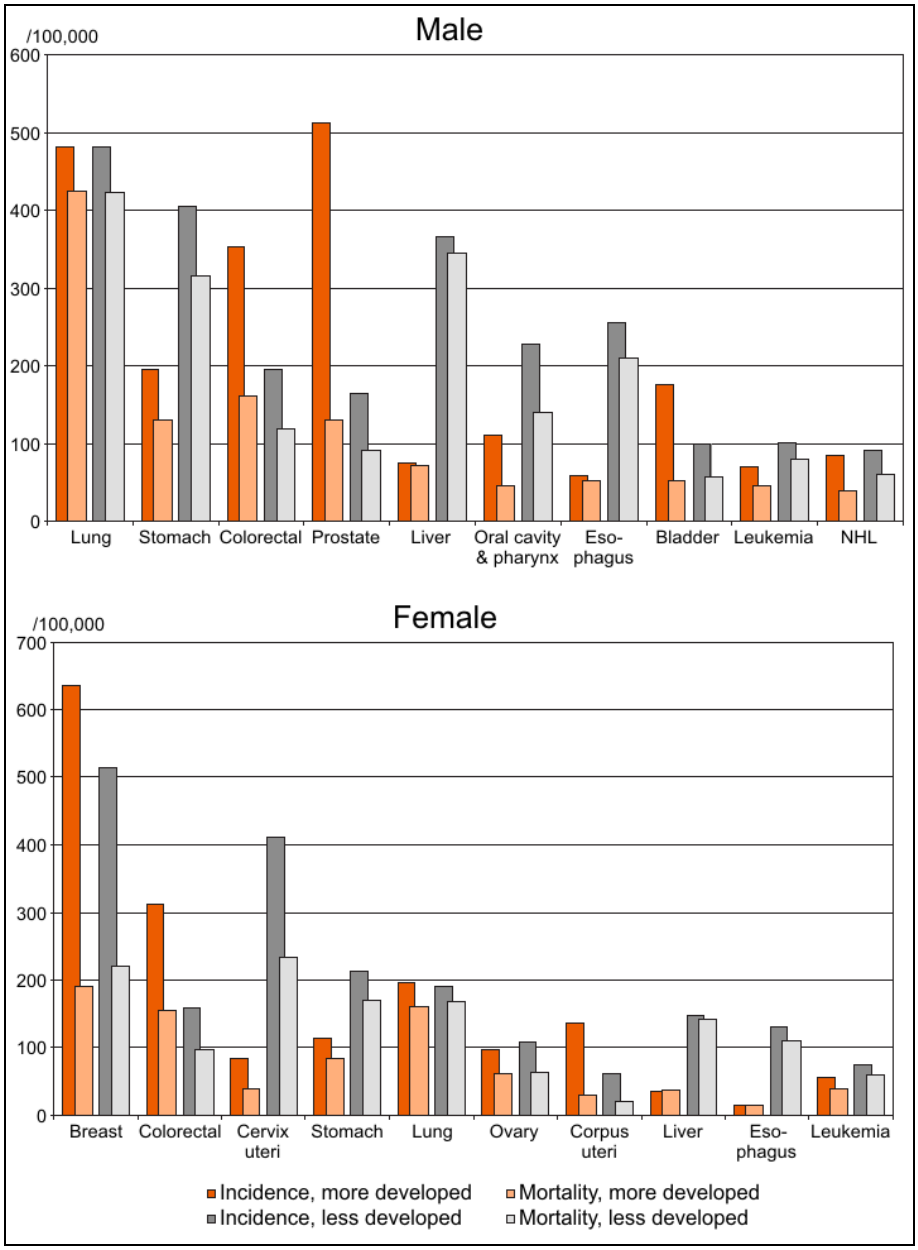


Figure 1: Incidence and mortality rates of the most common cancers in more-developed and less-developed countries worldwide.

Adopted from the IARC World Cancer Report 2008 (IARC 2008).

IARC – International Agency for Research on Cancer; NHL – Non-Hodgkin lymphoma

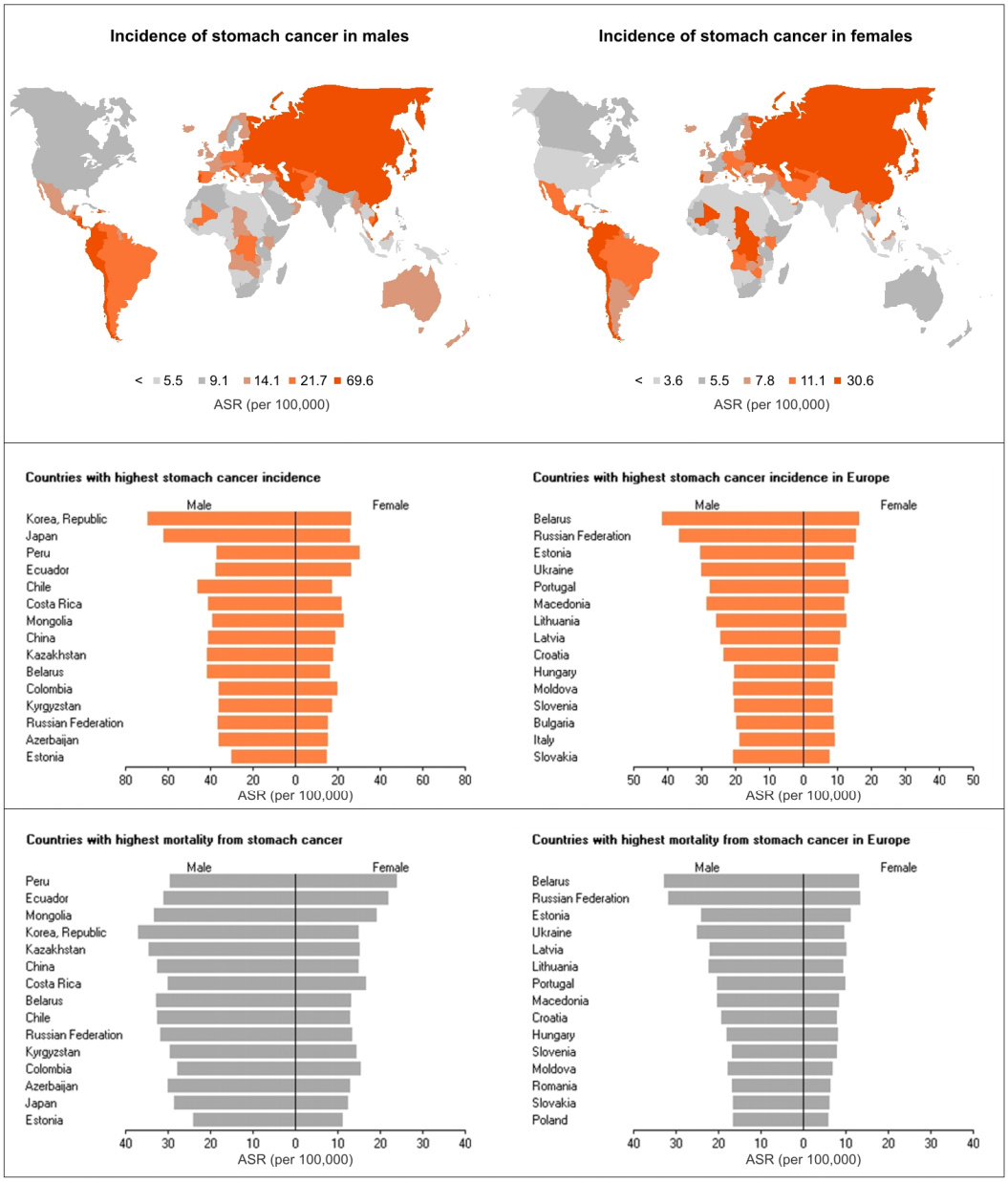


Figure 2: Incidence and mortality rates of gastric cancer worldwide and across Europe.
All figures were generated using the GLOBOCAN 2002 database (Ferlay *et al*. 2004).
ASR – age-standardized ratio (all ages included)

In Germany, gastric cancer is the fifth most common cancer among males and the seventh most common among females. Cumulative relative five-year survival rates are currently 35% for males and 31% for females. These survival rates are explicitly low if compared to other cancer types with high incidence. For example,

the five-year survival rate for colorectal cancer is about 60%, for breast cancer about 81% and for prostate cancer about 87% in Germany at the moment (Robert Koch Institut und Gesellschaft der epidemiologischen Krebsregister in Deutschland e.V. 2008).

1.2 Histopathology and histological classification

Cancers of the stomach include adenocarcinomas, adenosquamous carcinomas, squamous cell carcinomas, small cell carcinomas, gastric parietal cell carcinomas, undifferentiated carcinomas, endocrine cell tumors and other very rare types (Day *et al*. 2003, Hamilton and Aaltonen 2000, Watanabe *et al*. 1990). However, adenocarcinomas comprise ~95% of the total number of gastric malignancies (Schwartz 1996). According to histological growth pattern, gastric adenocarcinomas can be divided into two major subtypes, the "intestinal" and "diffuse" type. This classification system was established by the Finnish pathologist Pekka A. Laurén in 1965 (Lauren 1965) and proven to be useful by many later studies. Histologically, the intestinal type is characterized by cohesive tumors cells that form gland-like structures with lumina and grow tissue expanding (Lauren 1965, Ming 1977). Its tumor cells are rather large, usually fairly well polarized, columnar cells that can be easily recognized (Day *et al*. 2003). In the diffuse type, tumor cells loose cell to cell interaction and infiltrate the stroma of the stomach as single cells or small subgroups, causing a microscopical picture of non-cohesive, scattered tumor cells (Lauren 1965, Ming 1977). Gland formation is inconspicuous, except sometimes in the superficial part of the tumor (Day *et al*. 2003). Diffuse tumors often produce vast extends of mucus. The mucus can either be secreted to the extracellular space or remain within the cytosol (intracytoplasmic mucus) of tumor cells. If the mucus is not secreted and remains within the cells, the nucleus is occasionally pushed to the side, leading to a signet ring-like picture of tumor cells. If these so-called signet ring cells dominate the histological appearance, the tumor is

termed a signet ring cell carcinoma (Hamilton and Aaltonen 2000). Tumors that consist to more than 50% of extracellular mucus are referred to as mucinous (Hamilton and Aaltonen 2000). Besides strong production of mucus, most diffuse tumors are accompanied by excessive deposition of extracellular matrix (*e.g.* collagen) within the stroma, a phenomenon called fibrosis or desmoplasia. It is caused by activation of fibroblasts and leads to a general thickening of the stomach wall. Such tumors are often referred to as scirrhous gastric carcinomas. If tumor cells and accompanying fibrosis have spread across the whole stomach wall, the organ may become so constricted, inelastic and rigid that it resembles a leather bottle. Therefore, this special kind of diffuse-type tumor is termed "*linitis plastica*". In general, proliferation of connective tissue and mucus production is less prominent in intestinal-type tumors (Day *et al*. 2003, Remmele 1996). Microscopical pictures of representative samples of diffuse and intestinal-type adenocarcinomas are shown in Figure 3. Mixed types of both growth patterns and unclassified tumors exist as well, but are less frequent (Lauren 1965).

Although the Laurén's classification system dates back to 1965, it is still widely accepted and used by pathologists and surgeons in our days and represents a simple and robust classification approach. Other, more complex classification systems have been established over the years, certainly. However, all of them essentially overlap with the Laurén system (Fig. 4). For example, the well and moderately differentiated "tubular" and "papillary" type as defined by the WHO (Hamilton and Aaltonen 2000, Watanabe *et al*. 1990) and the "expanding" type according to Ming's classification (Ming 1977) correspond roughly to the intestinal type. In contrast, the WHO types "undifferentiated carcinoma" and "signet ring cell carcinoma" as well as the "infiltrative" type of Ming's classification generally match the diffuse type (Roukos *et al*. 2002, Vauhkonen *et al*. 2006, Werner *et al*. 2001).

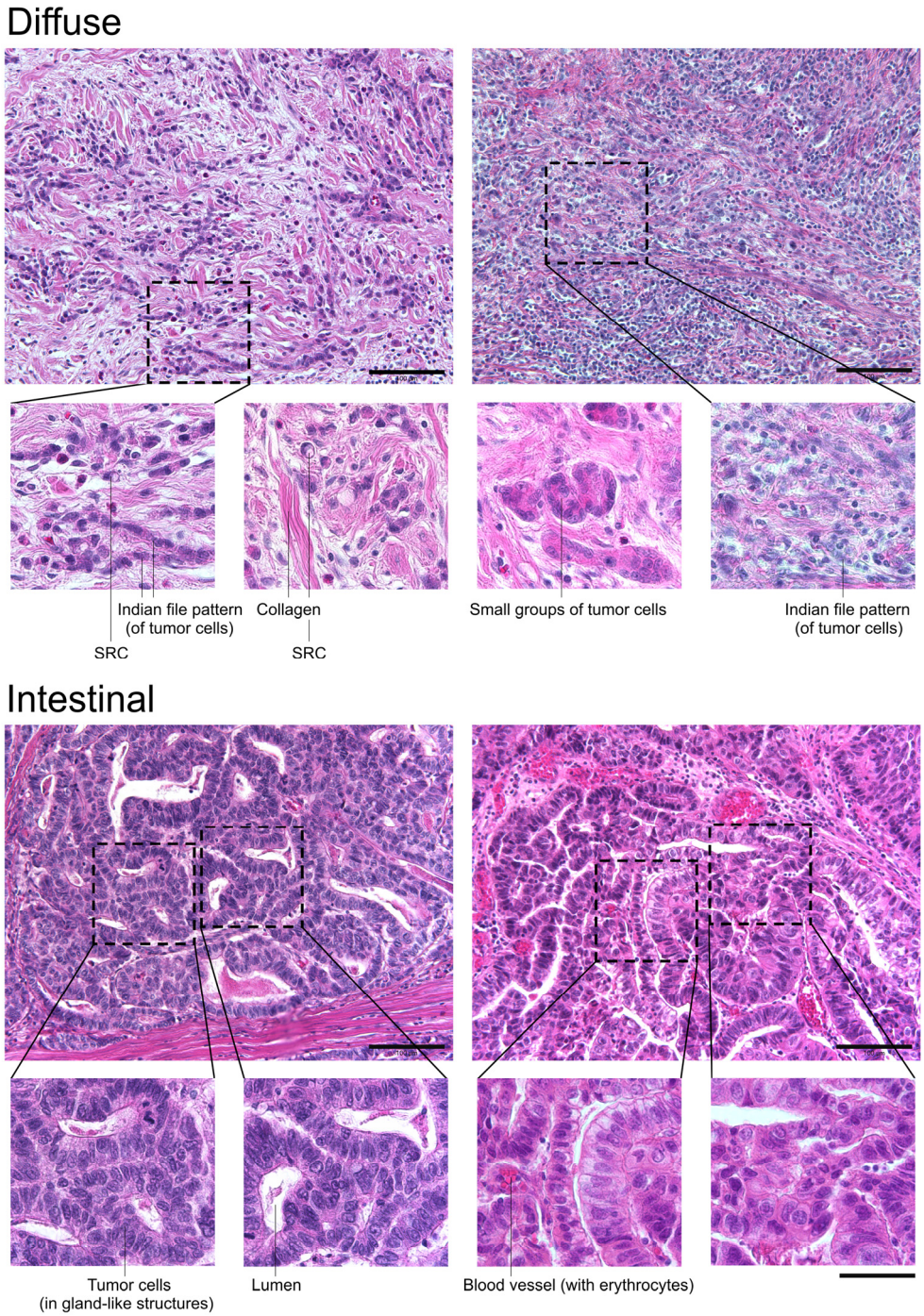


Figure 3: Microscopical view of human diffuse and intestinal-type gastric adenocarcinomas.

5 µm thin sections of formalin-fixed and paraffin-embedded tissues were HE-stained. Overview pictures are 200× magnified (scale bar represents 100 µm) and zoom-in areas are 400× magnified (scale bar represents 50µm). Representative sections are depicted.

SRC – signet ring cell; HE – hematoxylin-eosin

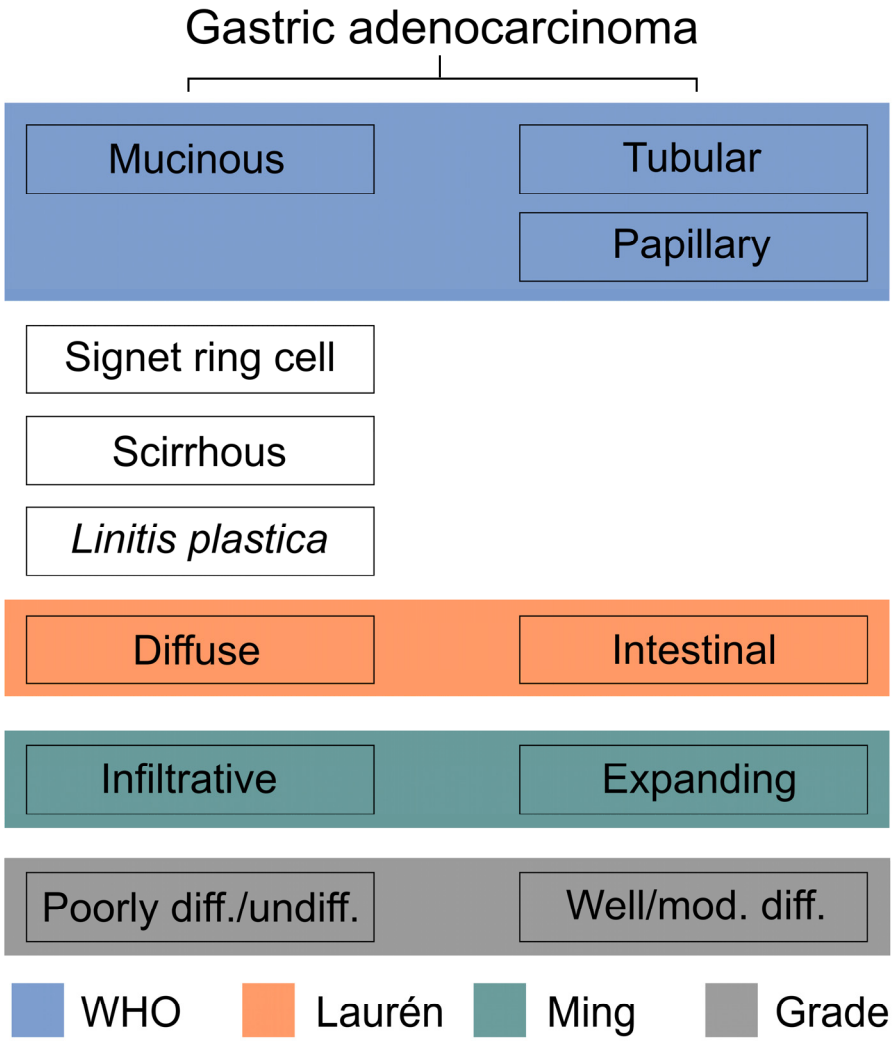


Figure 4: Major histological classification systems of human gastric adenocarcinomas and their overlap.

The major histological types as defined by the classification systems of the WHO, Laurén and Ming, plus additional special forms (white boxes), are illustrated in concordance to their predominant grade of differentiation. In principle, mucinous, tubular and papillary types may be graded as well, moderately or poorly differentiated (Day *et al*. 2003). However, the majority of mucinous tumors are poorly/undifferentiated, whereas the majority of tubular and papillary tumors are well to moderately differentiated, as depicted in this scheme. Other more infrequently used histological classification systems like the ones of Goseki (Goseki *et al*. 1992), Carneiro (Carneiro 1997) and Mulligan and Rember (Mulligan 1972) are excluded. Nevertheless, all of them roughly correspond to the Laurén's system as well (Day *et al*. 2003).

diff. – differentiated; mod. – moderately; WHO – World Health Organization; grade – histopathological grade (of differentiation)

In principle, diffuse-type tumors are explicitly more invasive and aggressive than intestinal ones, and affected patients possess extremely poor prognosis (Adachi *et al*. 2000b, Davessar *et al*. 1990, Ribeiro *et al*. 1981, Stemmermann and Brown 1974).

Intestinal-type tumors account for the majority of cancers. However, the ratio of intestinal to diffuse type has been declining over the last decades. The incidence of intestinal-type tumors has steadily decreased, accounting for the general decline of gastric cancer incidence, whereas the incidence of diffuse-type tumors, especially that of signet ring cell carcinomas, has slightly increased (Henson *et al*. 2004, Lauren and Nevalainen 1993, Munoz and Connelly 1971).

1.3 Etiology, epidemiology and carcinogenesis

Diffuse and intestinal-type gastric adenocarcinomas appear to have a distinctive etiology, epidemiology and follow different carcinogenesis and progression pathways (Ganten and Ruckpaul 1998). The intestinal type is referred to as the "epidemic form", because its overriding etiological factors are of environmental nature and related to diet and/or infection (*Helicobacter pylori*) (Correa and Shiao 1994, Ganten and Ruckpaul 1998). It is the most common variant in high-risk populations and accounts for the high gastric cancer incidence observed there. This type is preceded by a sequential chain of well-characterized events, of which chronic active gastritis, atrophy, intestinal metaplasia of the small bowel and colonic type, intraepithelial neoplasia and adenocarcinoma are the main stages (Correa 1988). However, its carcinogenic process does not necessarily need to include all these stages. Even *de novo* development is possible (Tahara 2004). Furthermore, some authors assume that atrophy and metaplasia are paracancerous rather than precancerous lesions (Hattori 1986, Meining *et al*. 2001).

The diffuse type, in contrast, is regarded as the "endemic form" having a rather strong hereditary/genetic bias and thus a similar frequency in areas of low and high incidence. The role of environmental factors for this type appears to be less important. It lacks well-recognizable precursor lesions. Hence, the existence of a multistep carcinogenic pathway is questionable.

Intestinal-type adenocarcinomas usually occur at late ages and predominate in males, whereas diffuse-type ones commonly arise in younger people, with males and females being affected in equal rates.

1.4 Molecular biology

The molecular biology varies substantially between diffuse and intestinal-type gastric adenocarcinomas. Multiple molecular characteristics have been identified to differ including mRNA and/or protein expression profile, gene copy numbers, microsattelite instability, loss of heterozygosity and mutation profile (Vauhkonen *et al*. 2006).

High-throughput gene expression profiling had identified multiple genes with differences in mRNA expression between the two types (Boussioutas *et al*. 2003, Hippo *et al*. 2002, Jinawath *et al*. 2004, Wu *et al*. 2006). However, little overlap of published gene lists exists forcing the demand for further and more comprehensive analyses. Moreover, the gene expression profile of the two types has not yet been analyzed on a genome-wide level. This suggests that many differentially expressed genes or even strong marker or classifier genes remain undiscovered so far.

Due to these multiple molecular and clinical differences, diffuse and intestinal-type gastric adenocarcinomas are widely accepted to represent distinct disease entities that may benefit from different therapeutic approaches (Chan *et al*. 1999).

1.5 Staging and assessment of patient prognosis

As mentioned previously, the stage at which a tumor is detected is essential for a patient's prognosis and survival. Therefore, correct assessment of the disease stage is crucial for appropriate patient management. In general, malignant tumors are staged according to the TNM classification system (Tab. 1, for gastric adenocarcinomas). It was first developed by Pierre Denoix in 1943–1952 and is continuously improved by the *Union Internationale Contre Le Cancer* (UICC) since 1950. It is the most widely used tool for classifying how far a cancer has spread from its point of origin and serves as a global standard. The TNM staging system comprises the depth of penetration of the primary tumor into the surrounding healthy tissue (T-stage), the number of regional metastatic lymph nodes (N-stage) and the presence or absence of distant metastases (M-stage). All three parameters are combined in the UICC-stage, which therefore represents a centralized parameter of cancer stage (Fig. 5).

Table 1: The TNM staging system of human gastric adenocarcinomas.

Adopted from the 5th and 7th edition of "TNM Classification of Malignant Tumours" (Sobin and Wittekind 1997, Sobin *et al*. 2010). The 7th edition is depicted, because it represents the most current version. The 5th edition is additionally shown, because patients enrolled for this retrospective study were staged according to it. The 6th edition is excluded, since it only differs from the 5th edition in subdividing the T2-stage into T2a (invades muscularis propria) and T2b (invades subserosa) stage.

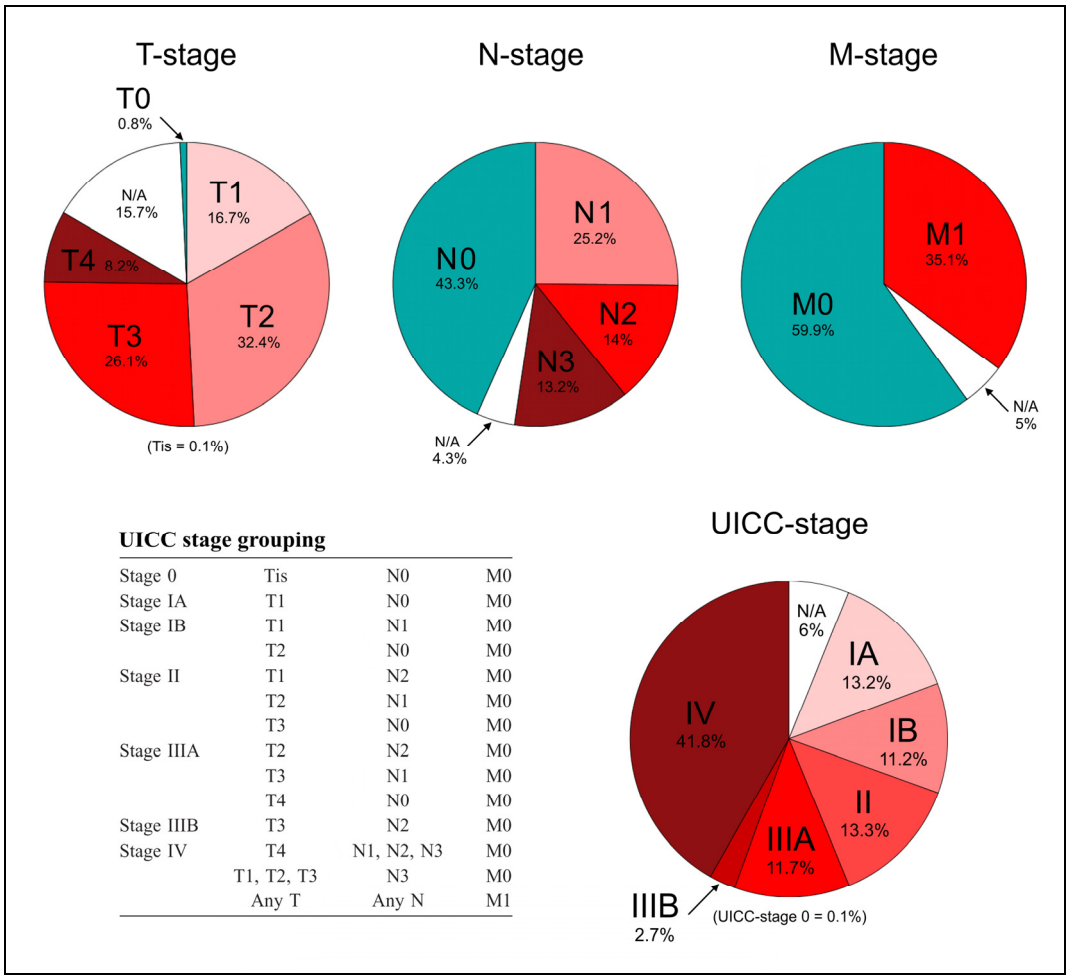
T – depth of penetration of primary tumor; N – metastatic involvement of regional lymph nodes; M – presence of distant metastases

5th edition (1997)		**7th edition (2010)**	
T = Primary tumor			
Tx	Cannot be assessed	Tx	Cannot be assessed
T0	No evidence of primary tumor	T0	No evidence of primary tumor
Tis	Carcinoma *in situ*	Tis	Carcinoma *in situ*
T1	Invades lamina propria/submucosa	T1a	Invades lamina propria
		T1b	Invades submucosa
T2	Invades muscularis propria/subserosa	T2	Invades muscularis propria
T3	Invades serosa (visceral peritoneum)	T3	Invades subserosa
T4	Invades adjacent structures	T4a	Invades serosa
		T4b	Invades adjacent structures
N = Regional lymph nodes			
Nx	Cannot be assessed	Nx	Cannot be assessed
N0	No regional lymph nodes are involved	N0	No regional lymph nodes are involved
N1	Metastases in 1–6 nodes	N1	Metastases in 1–2 nodes
N2	Metastases in 7–15 nodes	N2	Metastases in 3–6 nodes
N3	Metastases in >15 nodes	N3a	Metastases in 7–15 nodes
		N3b	Metastases in >15 nodes
M= Distant metastases			
Mx	Cannot be assessed		
M0	No distant metastases	M0	No distant metastases
M1	Distant metastases	M1	Distant metastases

The N-stage and M-stage are somewhat dependent on the T-stage, at least in the majority of cases, and increase with advancing T-stages. For example, nodal involvement is only seen in ~9–15% of Tis or T1-stage ("early" gastric cancer), whereas in T2–T4-stage up to 70% of patients have evidence of metastatic spread to the regional lymph nodes (Maruyama *et al*. 1989, McLean 2004).

As for most cancers, the TNM system is currently considered to be the most reliable prognostic factor for gastric adenocarcinoma patients (Adachi *et al*. 2000a, Msika *et al*. 2000, Siewert *et al*. 1998, Yokota *et al*. 2000, Yokota *et al*. 2004). Beyond TNM grading, other clinicopathological features like histological type (Adachi *et al*. 2000b, Davessar *et al*. 1990, Ribeiro *et al*. 1981, Stemmermann and Brown 1974), size and location of tumor (tumor site), macroscopic type of growth, venous invasion, lymphatic invasion (Msika *et al*. 2000, Yokota *et al*. 2004) and microvessel count as a measure of angiogenesis (Erenoglu *et al*. 2000) have been identified to comprise prognostic value. Prognostic value of age, gender and histopathological grading of differentiation (G) remains controversial and to be fully elucidated (Dicken *et al*. 2005). Beyond clinicopathological criteria, several molecules have been identified to be of prognostic value. These molecules essentially encompass general growth factors (*e.g.* EGF, TGFA/TGF-α) and their receptors (*e.g.* EGFR/ERBB1, ERBB2/HER-2, HGFR/c-met, FGFR2/K-sam), angiogenic growth factors (*e.g.* VEGF, TYMP/PDECGF, FGF2) and their receptors (*e.g.* VEGFR), angiogenic cytokines (*e.g.* IL8), cell cycle regulators (*e.g.* CDKN1B/p27, TP53/p53), cell adhesion molecules (*e.g.* CDH1/E-cadherin, CD44) and matrix-degrading enzymes (*e.g.* MMP1, MMP2, TIMP1) (Yasui *et al*. 2005).

Gastric adenocarcinoma is difficult to cure if not diagnosed at an early stage. Since early stage diseases are mostly accompanied by few symptoms, only, most tumors are advanced by the time of diagnosis (Dicken *et al*. 2005, Hundahl *et al*. 2000, Siewert *et al*. 1998). For instance, more than 65% of all gastric adenocarcinoma patients enrolled in Berlin's Robert-Rössle-Hospital between 1992 and 2007 had been diagnosed with tumors that had already invaded into the muscularis propria or deeper (T2–T4), whereas less than 20% had "early" disease being restricted to mucosa and submucosa (Tis or T1). More than half of all patients possessed involvement of regional lymph nodes and a strikingly large number of 42% had diagnosed UICC-stage IV (Fig. 5).



UICC stage grouping

Stage 0	Tis	N0	M0
Stage IA	T1	N0	M0
Stage IB	T1	N1	M0
	T2	N0	M0
Stage II	T1	N2	M0
	T2	N1	M0
	T3	N0	M0
Stage IIIA	T2	N2	M0
	T3	N1	M0
	T4	N0	M0
Stage IIIB	T3	N2	M0
Stage IV	T4	N1, N2, N3	M0
	T1, T2, T3	N3	M0
	Any T	Any N	M1

Figure 5: Distribution of different T, N, M and UICC-stages at gastric adenocarcinoma diagnosis in Germany (exemplified by patients of Robert-Rössle-Hospital, Germany).

Unpublished data of patients, who had undergone resection for gastric adenocarcinoma at Robert-Rössle-Hospital between 1992 and 2007 (~780 patients) was used for creating pie charts. TNM staging and UICC stage grouping had been performed according to guidelines of the 5th edition of "TNM Classification of Malignant Tumours" (Sobin and Wittekind 1997).

T – depth of penetration of primary tumor; N – metastatic involvement of regional lymph nodes; M – presence of distant metastases; UICC – *Union Internationale Contre Le Cancer*; N/A – data not available

Complete surgical resection (no residual tumor left; R0) is the major goal of gastric adenocarcinoma treatment, because it is the basic prerequisite for curing this disease. However, only few patients with apparent R0 resection are fully cured. Most of them (~70%) are faced with recurrence of disease (*e.g.* metastases, loco-regional recurrence) during the following years, leading to shortened survival (Macdonald *et al*. 2001, McLean 2004).

Naturally, complete resection becomes more difficult and improbable with advancing stages. However, the majority (~70%) of gastric tumors can be resected completely today (Siewert *et al*. 1998), which is mainly accomplished by improvement of surgical techniques. For these completely resected patients, the N-stage is the most important prognostic parameters to predict clinical outcome and survival (Siewert *et al*. 1998). Patients with tumors accompanied by lymph node metastases at diagnosis clearly show a decreased survival rate when compared to lymph node-negative patients (Hohenberger and Gretschel 2003, Kim *et al*. 2006, Lee *et al*. 2007, Maehara *et al*. 2002, Siewert *et al*. 1998). Therefore, exact assessment of the extent of nodal spread is crucial for predicting patient prognosis and tailoring choice of therapy. Estimation of the stage of nodal involvement is challenging and depends on the number of lymph nodes examined, the extent of lymph node dissection (perigastric or perigastric plus extraperigastric nodes) and, of course, the experience of the surgeon. Thus, the question of which and how many lymph nodes should be dissected (type of lymphadenectomy) to allow accurate staging and adequate surgical treatment is subject to a long debate and still under discussion (Jansen *et al*. 2005, McCulloch *et al*. 2005). In addition, preoperative imaging techniques, such as ultrasound endoscopy or computed tomography, and intraoperative procedures, such as the sentinel node technique, are insufficient in reliably predicting nodal spread (Ajisaka and Miwa 2003, Jansen *et al*. 2005, Kim *et al*. 2005, Kim *et al*. 2004, Ryu *et al*. 2003).

Hence, numerous studies were devoted to the prediction of nodal status according to the molecular features of the primary tumor and were able to identify several marker or classifier genes/proteins for nodal involvement. Examples include BIK (BCL2-interacting killer, apoptosis-inducing), AURKB (aurora kinase B), EIF5A2 (eukaryotic translation initiation factor 5A2), CDH2 (cadherin 2, type 1, N-cadherin, neuronal), TRIP10 (thyroid hormone receptor interactor 10), SERPINB5 (serpin peptidase inhibitor, clade B, member 5), CEACAM5 & 6 (carcinoembryonic antigen-related cell adhesion molecule 5 & 6), SPARC (secreted protein, acidic, cysteine-rich) and S100A11; all being initially identified by gene

expression profiling studies (Hasegawa *et al*. 2002, Marchet *et al*. 2007, Mori *et al*. 2004, Norsett *et al*. 2004, Terashima *et al*. 2005, Wang *et al*. 2004).

1.6 Lymph node-positive patients

Metastatic dissemination to regional lymph nodes frequently accompanies gastric adenocarcinoma diagnosis. Its incidence increases with deeper invasion of the primary tumor into the gastric wall (T-stage) (Maruyama *et al*. 1989) and occurs with equal frequency in the different histological types (Day *et al*. 2003). Although most regional lymph node-positive (N+) patients show a decreased survival rate, a small subgroup of patients with clearly favorable prognosis exists (Kim *et al*. 2007a). Kaplan-Meier survival analysis of patient data from Robert-Rössle-Hospital reveals that patients with survival longer than 10 years can be identified among all three N+ stages (Fig. 6A).

Controversial observations were made to predict these differences in prognosis with clinicopathological parameters (Cheong *et al*. 2006, Kim *et al*. 2007a) and yet no unambiguous "clinicopathological answer" is available. Recently, efforts were made to distinguish these patient subgroups with different prognosis on the basis of molecular features of the primary tumor. In 2006, Jüttner *et al.* were able to identify VEGF-D and its receptor VEGFR-3, two important players of tumor-related lymphangiogenesis, to be differentially expressed between these prognostic patient subgroups. These molecules were detected to act as novel prognostic marker molecules (Juttner *et al*. 2006) that explain some of the differences in survival of N+ patients (Fig. 6B). These data first indicated that lymph node-positive gastric adenocarcinoma patients display differences in their molecular behavior, which are reflected on the prognostic level.

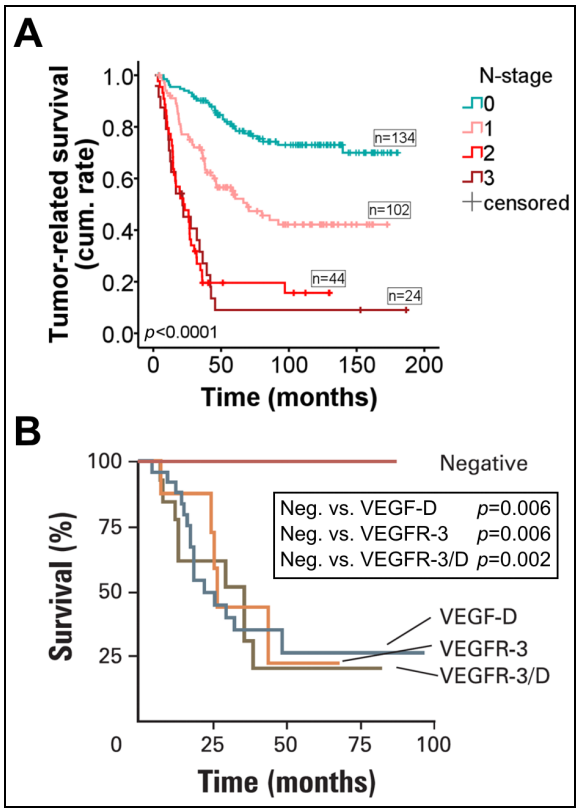


Figure 6: Survival of R0 M0 N+ gastric adenocarcinoma patients.

A: Tumor-related survival of patients with different N+ stages compared to N0 was assessed using the Kaplan-Meier method. In general, N+ patients possess reduced survival rates compared to N0 patients, but, within all three N+ stages, patient subgroups with survival longer than 10 years can be identified.
B: Tumor-related survival according to the presence or absence of VEGFD and its receptor VEGFR3 independent of N-stage and other clinical parameters (Juttner et al 2006). "Negative" refers to absence of both molecules.
Survival was assessed using the Kaplan-Meier method. *p*-values were calculated by logrank test.

N – metastatic involvement of regional lymph nodes; N+ – regional lymph node-positive; N0 – no regional lymph nodes involved; R0 – no residual tumor/complete resection; M0 – no distant metastases; VEGF – vascular endothelial growth factor; R – receptor; neg. – negative

1.7 Objectives of this thesis

The first major aim of this thesis was to further examine the molecular background of diffuse and intestinal-type gastric adenocarcinomas. Therefore, genome-wide expression profiles of both histological types should be established on the basis of a regional lymph node-positive patient cohort. Biological interpretation of these profiles should give insight into the molecular mechanisms underlying carcinogenesis and progression, and the biological behavior of either type. However, the diffuse type was of particular interest, because comparably little is known about its underlying molecular mechanisms and its eminent invasiveness. Advances in research on this topic may lead to the identification of novel targets for treatment allowing a more targeted and personalized therapy in future.

The second major aim was to establish a gene expression profile-based stratification of regional lymph node-positive gastric adenocarcinoma patients with respect to prognosis. In association with that, gene signatures and/or single marker genes correlating with clinical outcome and prognosis of these patients should be identified. Such prognostic gene signatures or marker genes may allow more accurate prediction of prognosis of node-positive gastric adenocarcinoma patients and thus selection of more tailored and personalized therapy, with less patients being over or undertreated.

Regional lymph node-positive patients were chosen as the cohort of this study, because involvement of regional lymph nodes frequently accompanies gastric adenocarcinoma diagnosis. In Germany, for example, these patients make up more than half (*e.g.* ~70% in Siewert *et al.*, 1998; or ~52% in unpublished data of Robert-Rössle-Hospitel, refer to Fig. 5) of the patients diagnosed with gastric adenocarcinoma. Thus, they are a very representative patient population of this disease entity. Furthermore, especially little is known about the molecular mechanisms underlying progression and recurrence of these, initially disseminated, tumors after complete resection. Previous studies aiming at molecular stratification of gastric cancer patients with regard to disease outcome and at the identification of prognostic markers were performed on node-negative patients or mixed cohorts, in fact (Chen *et al.* 2005, Chen *et al.* 2003, Lee *et al.* 2010, Leung *et al.* 2002, Leung *et al.* 2004, Takeno *et al.* 2010).

2 Material and methods

2.1 Material

2.1.1 Gastric adenocarcinoma patients and tumor samples

Tumor tissue samples used for this study were obtained from 65 patients of the Robert-Rössle-Hospital. All patients had been diagnosed with regional lymph node-positive gastric adenocarcinoma and had undergone resection of tumor with curative intention between 1995 and 2003 in this hospital.

Resected tumors had been classified and staged according to the WHO classification (Hamilton and Aaltonen 2000, Watanabe *et al.* 1990) and TNM staging system (Sobin and Wittekind 1997). Additional histopathological parameters that had been assessed included histological type according to Laurén (Lauren 1965), histopathological grading of differentiation (G), venous invasion (V) and lymphatic invasion (L).

After surgery, all patients had entered into a regular follow-up program, which included physical examination, laboratory tests (for tumor markers CEA, CA19-9, CA72-4), transabdominal ultrasonography, chest radiography and computed tomography. Follow-up of patients with non-recurrent disease ("cured" patients) was 3.6 to 12.7 years with a median of 6.5 years. Recurrent patients were followed until progression or death.

In order to identify prognostic gene signatures or marker genes, it is crucial to avoid as many systematic errors as possible and to keep the patient population as homogeneous as possible. To do so, patients were split into two independent his-

tological cohorts, a diffuse and an intestinal-type cohort (according to Laurén's classification), and only patients that had been distant metastasis-free by the time of diagnosis and surgery (M0), that had no secondary malignancy and no residual tumor after surgery (R0) were included. Further criteria which needed to be fulfilled included "appearance of first recurrence at least 2 months after surgery", "absence of postoperative death" and "no treatment with neoadjuvant or adjuvant chemotherapy". Patients with N3-stage were also excluded, though only from the cohort of intestinal-type tumors. An exclusion of N3-stage patients from the diffuse-type cohort was not performed, due to lack of sufficient amounts of diffuse-type tumor samples. However, only two patients from this cohort had diagnosed N3-involvement. The reason for removal of N3-stage patients is that metastatic dissemination to the lymph nodes is so advanced in this stage that a clear separation from distant metastases is not possible anymore. Diagnosed N3-stage gives only information about the number of involved nodes (>15), but not about the region where the metastatic nodes are situated. Detailed information on clinicopathological parameters and progression-based parameters of the patient cohorts can be found in Table 2 and 3.

Tumor samples were snap frozen immediately after resection and pathologic survey and stored constantly at –80 °C. For total RNA extraction, tumor samples were cut with a cryostat in 5 µm thin cryosections. The first and the last section of each sample were HE-stained and examined again by a pathologist for tumor content and histological type. The remaining cryosections were used for RNA extraction. Only samples with tumor contents higher than 50% were included in microarray analysis, namely 59.

Table 2: Clinicopathological data of 65 N+ gastric adenocarcinoma patients enrolled for the identification of prognostic gene signatures.

Identification of prognostic genes was performed separately for diffuse and intestinal-type patients. Hence, data is presented for both histological cohorts, respectively. TNM/UICC staging was performed according to 5th edition (1997) of "TNM Classification of Malignant Tumours" (Sobin and Wittekind 1997).

T – depth of penetration of primary tumor; N – metastatic involvement of regional lymph nodes; M – presence of distant metastases; UICC – *Union Internationale Contre Le Cancer*; G – grade (of differentiation); L – lymphatic invasion; V – venous invasion; R0 – no residual tumor/complete resection; N/A – data not available; n – number

	Intestinal patients (n=36)		**Diffuse patients (n=29)**	
Clinicopathological parameter	**n**	**%**	**n**	**%**
Sex				
Male	23	63.9	16	55.2
Female	13	36.1	13	44.8
Age				
35–50	5	13.9	7	24.1
51–70	18	50.0	21	72.4
71–86	13	36.1	1	3.5
Tumor location				
Upper third of stomach	3	8.3	2	6.9
Middle third of stomach	7	19.4	9	31.0
Lower third of stomach	9	25.0	10	34.5
Cardia	17	47.2	5	17.2
Total stomach involved	0	0	3	10.3
T-stage				
T1	2	5.6	1	3.4
T2	19	52.8	17	58.6
T3	13	36.1	9	31.0
T4	2	5.6	2	6.9
N-stage				
N1	25	69.4	17	58.6
N2	11	30.6	10	34.5
N3	0	0	2	6.9
M-stage				
M0	36	100	29	100
UICC-stage				
IB	2	5.6	1	3.4
II	11	30.6	10	34.5
IIIA	18	50.0	11	37.9
IIIB	3	8.3	3	10.3
IV	2	5.6	4	13.8
G-stage				
G2	8	22.2	2	6.9
G3	28	77.8	27	93.1
L-stage				
No	15	41.7	14	48.3
Yes	20	55.6	15	51.7
N/A	1	2.8	0	0

table continues

continued

	Intestinal patients (n=36)		Diffuse patients (n=29)	
Clinicopathological parameter	**n**	**%**	**n**	**%**
V-stage				
No	22	61.1	18	62.1
Yes	12	33.3	11	37.9
N/A	2	5.6	0	0
Neoadjuvant chemotherapy	0	0	0	0
Adjuvant chemotherapy	0	0	0	0
Secondary carcinoma	0	0	0	0
R0 resection	36	100	29	100
Curative treatment intention	36	100	29	100

Table 3: Disease progression-based data of 65 N+ gastric adenocarcinoma patients enrolled for the identification of prognostic gene signatures.

Identification of prognostic gene signatures was performed separately for diffuse and intestinal-type tumors. Hence, data is presented for both histological cohorts, respectively.

n – number

	Intestinal patients (n=36)		Diffuse patients (n=29)	
Progression-based parameter	**n**	**%**	**n**	**%**
Postoperative recurrence of disease				
Yes	29	80.6	17	58.6
No	7	19.4	12	41.4
Postoperative metastases				
Yes	26	72.2	17	58.6
No	10	27.8	12	41.4
Hematogenous metastases				
Yes	19	52.8	13	44.8
No	17	47.2	16	55.2
Lymphatic metastases				
Yes	11	30.6	8	27.6
No	25	69.4	21	72.4
Peritoneal metastases				
Yes	9	25.0	11	37.9
No	27	75.0	18	62.1
Locoregional recurrence				
Yes	6	16.7	7	24.1
No	30	83.3	22	75.9

2.1.2 General chemicals and reagents

Acetic acid	Merck KGaA, Darmstadt, GER
Acetone (≥99.8%, p.a. ACS, ISO)	Carl Roth GmbH + Co. KG, Karlsruhe, GER
3-amino-9-ethylcarbazole (AEC) tablets	Sigma-Aldrich Corporation, St. Louis, USA
Blocking reagent	Boehringer Ingelheim GmbH, Ingelheim, GER
BM Purple	Roche Diagnostics GmbH, Mannheim, GER
Chloroform (100%, p.a.)	Carl Roth GmbH + Co. KG, Karlsruhe, GER
Citric acid (≥99.5%, p.a. ACS reagent)	Sigma-Aldrich Corporation, St. Louis, USA
DNA, single-stranded from salmon testes	Sigma-Aldrich Corporation, St. Louis, USA
Dextran sulfate (from dextran with average MW=500kg/mol)	Sigma-Aldrich Corporation, St. Louis, USA
4',6-diamidino-2-phenylindole-dihydrochloride (DAPI)	Roche Diagnostics GmbH, Mannheim, GER
Entellan® mounting medium	Merck KGaA, Darmstadt, GER
Eosin G (yellowish)	Merck KGaA, Darmstadt, GER
Ethanol (≥99.8%, p.a.)	Carl Roth GmbH + Co. KG, Karlsruhe, GER
Ethylenediaminetetraacetic acid (EDTA)-dihydrate	Sigma-Aldrich Corporation, St. Louis, USA
Fetal Bovine Serum, qualified, heat-inactivated	Invitrogen GmbH, Karlsruhe, GER
Fluorescence mounting medium	Dako, Glostrup, DEN
Glycine	Carl Roth GmbH + Co. KG, Karlsruhe, GER

Hematoxylin (Mayer's) ("*Hämalaulösung sauer nach Mayer*")	Carl Roth GmbH + Co. KG, Karlsruhe, GER
Heparin sodium salt from porcine intestinal mucosa (≥140 USP units/mg)	Sigma-Aldrich Corporation, St. Louis, USA
Hydrogen peroxide 30% (H_2O_2)	Merck KGaA, Darmstadt, GER
Nuclear Fast Red (Certistain®)	Merck KGaA, Darmstadt, GER
Levamisole Endogenous Alkaline Phosphatase Inhibitor	Dako, Glostrup, DEN
Magnesium chloride hexahydrate ($MgCl_2(H_2O)_6$)	Sigma-Aldrich Corporation, St. Louis, USA
dNTPs	MBI Fermentas GmbH, St. Leon Rot, GER
Paraformaldehyde	Sigma-Aldrich Corporation, St. Louis, USA
Pertex® mounting medium	MEDITE GmbH, Burgdorf, GER
2-Propanol	Merck KGaA, Darmstadt, GER
RNase AWAY	Invitrogen GmbH, Karlsruhe, GER
Sheep serum (normal)	Millipore Corporation, Billerica, USA
SIGMACOTE	Sigma-Aldrich Corporation, St. Louis, USA
Sodium acetate (CH_3COONa)	Merck KGaA, Darmstadt, GER
Sodium chloride ($NaCl_2$)	Merck KGaA, Darmstadt, GER
Sodium dodecyl sulfate	SERVA Electrophoresis GmbH, Heidelberg, GER
Sodium hydroxide (NaOH)	Merck KGaA, Darmstadt, GER
Tris	base SERVA Electrophoresis GmbH, Heidelberg, GER
Trisodium dihydrate	Merck KGaA, Darmstadt, GER

Triton X-100	Sigma-Aldrich Corporation, St. Louis, USA
tRNA from E. coli (MRE 600); RNase-negative	Roche Diagnostics GmbH, Mannheim, GER
TRIzol reagent	Life Technologies, Grand Island, USA
Tween 20	Carl Roth GmbH + Co. KG, Karlsruhe, GER
Xylene	Carl Roth GmbH + Co. KG, Karlsruhe, GER

2.1.3 Enzymes and pre-developed enzyme mixes

DNase I (RNase Free DNase Set)	Qiagen Inc., Valencia, USA
RNA UltraSense™ One-Step Quantitative RT-PCR System (contains SuperScript™ III reverse-transcriptase and Platinum *Taq* DNA polymerase)	Invitrogen GmbH, Karlsruhe, GER
SuperScript II Reverse Transcriptase	Invitrogen GmbH, Karlsruhe, GER
TaqMan® Universal PCR Mastermix, No AmpErase® UNG (contains Ampli*Taq* Gold® DNA polymerase)	Applied Biosystems, Life Technologies Corporation, Carlsbad, USA

2.1.4 Antibodies

2.1.4.1 Primary antibodies

Mouse monoclonal anti-human ACTA2 [1A4] antibody (Cat. #: M0851)	Dako, Glostrup, DEN
Sheep anti-Digoxigenin-AP (fab fragments)	Roche Diagnostics GmbH, Mannheim, GER

Mouse monoclonal anti-pan KRT [B311.1] antibody (Cat. #: GTX28474)	GeneTex Inc., Irvine, USA
Mouse monoclonal anti-human ProCOLI antibody (Cat. #: MAB1913)	Millipore Corporation, Billerica, USA
Goat polyclonal anti-human THBS4 antibody (Cat. #: AF2390)	R&D Systems, Minneapolis, USA
Mouse monoclonal anti-human THBS4 antibody (Cat. #: MAB2390)	R&D Systems, Minneapolis, USA
Mouse monoclonal anti-VIM [13.2] antibody (Cat. #: V 5255)	Sigma-Aldrich Corporation, St. Louis, USA

2.1.4.2 Secondary antibodies

Donkey Alexa Fluor® 555 anti-goat IgG (H+L) (Cat. #: A21432)	Invitrogen GmbH, Karlsruhe, GER
Horse biotinylated anti-mouse IgG (H+L) (Cat. #: BA-2000)	Vector Laboratories Inc., Burlingame, USA
Donkey Alexa Fluor® 488 anti-mouse IgG (H+L) (Cat. #: A21202)	Invitrogen GmbH, Karlsruhe, GER
Goat biotinylated anti-rabbit IgG (H+L) (Cat. #: BA-1000)	Vector Laboratories Inc., Burlingame, USA

2.1.5 Kits

Biotin Blocking System	Dako, Glostrup, DEN
GeneChip® Eukaryotic Poly-A RNA Control Kit	Affymetrix Inc., Santa Clara, USA
GeneChip® Expression 3'-Amplification One-Cycle cDNA Synthesis Kit	Affymetrix Inc., Santa Clara, USA
GeneChip® Expression 3'-Amplification Reagents for IVT Labeling	Affymetrix Inc., Santa Clara, USA

GeneChip® Sample Cleanup Module	Affymetrix Inc., Santa Clara, USA
RNA 6000 Nano Kit	Agilent Technologies Inc. Santa Clara, USA
RNeasy Mini Kit	Qiagen Inc., Valencia, USA
RNeasy Micro Kit	Qiagen Inc., Valencia, USA
Vectastain Elite ABC Kit	Vector Laboratories Inc., Burlingame, USA

2.1.6 Cell culture media and solutions

DMEM, high glucose (4.5 g/l)	PAA Laboratories GmbH, Pasching, AUT
Fetal Bovine Serum "GOLD" EU approved	PAA Laboratories GmbH, Pasching, AUT
L-glutamine (200mM)	PAA Laboratories GmbH, Pasching, AUT
MEM Amino Acids	PAA Laboratories GmbH, Pasching, AUT
MEM Non Essential Amino Acids (NEAA)	PAA Laboratories GmbH, Pasching, AUT
PBS 10× (Dulbecco's PBS) without Ca & Mg	PAA Laboratories GmbH, Pasching, AUT
Penicillin/streptomycin (100×)	PAA Laboratories GmbH, Pasching, AUT
RPM1 1640 medium (without L-glutamine; with phenol red)	PAA Laboratories GmbH, Pasching, AUT
Sodium pyruvate (100mM)	PAA Laboratories GmbH, Pasching, AUT
Trypsin EDTA (1:250)	PAA Laboratories GmbH, Pasching, AUT

2.1.7 Solutions and buffers

Acetic buffer (pH 5.0)

7 parts 0.2M sodium acetate mixed with 3 parts 0.2M acetic acid

Blocking Solution (for *in situ* hybridization)

1% Boehringer blocking reagent

10% sheep serum

0.1% Triton X-100

in TBS buffer

Development solution (for immunohistochemistry)

1:20 AEC (3-amino-9-ethylcarbazole) solution (purchased tablets had been previously dissolved according to manufacturer's recommendations)

0.015% H_2O_2 (immediately added before usage)

in acetic buffer (pH 5.0)

Eosin

1% eosin

in 70% ethanol

Hybridization buffer

50% formamide

10% dextrane sulfate

5% Boehringer blocking reagent

0.1% Tween 20

100 µg/ml heparine

100 µg/ml tRNA

100 µg/ml ssDNA

5 mM EDTA

in 5× SSC

NTMT

50 mM $MgCl_2$

100 mM NaCl

100 MM Tris-HCl pH 9.5

0.1% Tween 20

in ddH_2O

4% paraformaldehyde (for *in situ* hybridization)

4% paraformaldehyde (dissolve in ddH_2O treated with NaOH at 60 °C while stirring)

1 mM $MgCl_2$ (add after paraformaldehyde is dissolved)

0.2 mM NaOH (add after paraformaldehyde is dissolved)

in 1× PBS

10× PBS (for immunohistochemistry and *in situ* hybridization)

10× Dulbecco's PBS (without Ca & Mg); ordered from PAA Laboratories GmbH

20× SSC

300 mM Natriumcitrat

3 M NaCl

in ddH_2O

First, one constituent is dissolved and then the other one. Finally, pH is adjusted to 7.0.

TBS buffer (for *in situ* hybridization)

150 mM NaCl

100 mM Tris-HCl pH 7.4

2 mM KCl

in ddH_2O

TES buffer (for *in situ* hybridization)

10 mM Tris-HCl pH 8.0

0.5 M NaCl

1 mM EDTA

in ddH_2O

2.1.8 Oligonucleotides (primer)

oligo(dT) primer (for cDNA synthesis) TTT TTT TTT TTT TTT T

2.1.9 Oligonucleotide (primer)-probe-mixes

All primer-probe-mixes were purchased from Applied Biosystems, Life Technologies Corporation, Carlsbad, USA as pre-developed and optimized assays (Tab. 4).

Table 4: Summary of primer-probe-mixes used in this thesis.

	Gene title	Gene symbol	"TaqMan® Gene Expression Assay"	Label
Endogenous control genes	Actin, *beta*	*ACTB*	4326315E	VIC/MGB
	Glyceraldehyde-3-phosphate dehydrogenase	*GAPDH*	4326317E	VIC/MGB
Genes of interest (target genes)	BMP and activin membrane-bound inhibitor homolog	*BAMBI*	Hs03044164_m1	FAM
	Ephrin receptor A4	*EPHA4*	Hs00953178_m1	FAM
	Folate receptor 1	*FOLR1*	Hs01124177_m1	FAM
	Growth associated protein 43	*GAP43*	Hs00967138_m1	FAM
	Homeobox C10	*HOXC10*	Hs00213579_m1	FAM
	RAN binding protein 17	*RANBP17*	Hs00224684_m1	FAM
	Ras-related associated with diabetes	*RRAD*	Hs00188163_m1	FAM
	Thrombospondin 4	*THBS4*	Hs00170261_m1	FAM

2.1.10 Consumable material

Cover slips 24×50 mm	Carl Roth GmbH & Co KG, Karlsruhe, GER
GeneChip® HG U133 Plus 2.0	Affymetrix Inc., Santa Clara, USA
QIAshredder mini spin column	Qiagen Inc., Valencia, USA
SuperFrost® plus microscope slides	Menzel GmbH & Co KG, Braunschweig, GER

2.1.11 Laboratory devices

ABI PRISM 7900 HT Sequence Detection System	Applied Biosystems, Life Technologies Corporation, Carlsbad, USA
2100 Bioanalyzer	Agilent Technologies, Santa Clara, USA
Camera RT KE Spot	Visitron Systems GmbH, Puchheim, GER
Confocal microscope Leica TCS SPE (on Leica DM2500)	Leica Microsystems GmbH, Wetzlar, GER
Cryostat HM 560 Cryo-Star	MICROM International GmbH, Walldorf, GER
Microscope Olympus BX51	Olympus Corporation, Tokyo, JPN
Mikro-Dismembrator U	B. Braun Biotech International GmbH, Melsungen, GER
NanoDrop ND-1000	NanoDrop Technologies, Wilmington, USA

2.1.12 Software and online services

Corel PHOTO-PAINT X3 13	Corel Corporation, Ottawa, CAN

EndNote 9	Thomson Reuters, New York, USA
GeneSpring GX 10.0.2	Agilent Technologies Inc. Santa Clara, USA
GOSSIP	MicroDiscovery GmbH, Berlin, GER
LAS AF	Leica Microsystems GmbH, Wetzlar, GER
MetaMorph 6.2r2	Molecular Devices, Downingtown, USA
Microsoft Office 2003	Microsoft Corporation, Redmond, USA
NetAffx Analysis Center	Affymetrix Inc., Santa Clara, USA
SDS 2.2	Applied Biosystems, Life Technologies Corporation, Carlsbad, USA
SPSS 16.0	SPSS Inc., Chicago, USA
ONCOMINE	https://www.oncomine.org (Rhodes *et al.* 2004)

2.2 Methods

2.2.1 Total RNA extraction from frozen tissue

A cryostat was used to cut frozen tissue samples. At the beginning of each tissue piece, a 5 µm thin section was mounted onto a microscope slide for later HE-staining and pathological evaluation of sample. Next, 10–30 µm thin sections were cut for RNA extraction, transferred into 1 ml of TRIzol reagent and disrupted using a Dismembrator with pre-cooled cups for 4–5 minutes. At last, a second 5 µm thin section was cut and mounted for HE-staining and pathological evaluation as well.

After disruption, the samples were incubated for 5 minutes at room temperature followed by addition of 0.2 ml chloroform. Next, samples were vortexed for ~15 seconds and incubated 2–3 minutes at room temperature. After centrifugation of samples at 12,000*g* for 15 minutes at 4 °C, the upper aqueous phase, which contains the RNA, was transferred to a new reaction tube. Equal volume of 70% ethanol (–20 °C cold) was added and mixed by vortexing. The obtained solution was transferred to an RNeasy isolation column and continuatively processed according to the RNeasy Mini Kit manufacturer's manual. During procedure, an on-column DNase I digest was performed according to manufacturer's recommendations to eliminate genomic DNA contamination.

2.2.2 Total RNA extraction from cells or cell pellets

Cells (living or pelleted) were lysed with RNeasy Lysis Buffer and disrupted using the QiaShredder according to the manufacturer's instructions. Total RNA extraction was performed using the RNeasy Mini or Micro Kit (depending on number of cells) with on-column DNase I digest according to manufacturer's manual.

2.2.3 Quantitation and quality assessment of RNA

RNA concentration and purity were determined spectrophotometrically using the NanoDrop ND-1000 as described elsewhere (Gallagher and Desjardins 2006).

Intact RNA is a prerequisite for a proper microarray study. Thus, integrity of extracted RNA was checked, using the Agilent Bioanalyzer 2100, before processing for microarray experiments. The Agilent Bioanalyzer performs a microfluidics-based electrophoretic separation of nucleic acids according to their size. For quality assessment of eukaryotic total RNA, the extent of degradation of the 18S and 28S ribosomal RNA peaks is measured and a so-called "RNA Integrity Number

(RIN)" is calculated. RINs from 1 to 10 can be assigned to a sample, with "perfectly" intact RNA obtaining a RIN of 10.

For this study, only RNAs possessing a RIN above 7 were regarded as satisfactory in quality and thus as suitable for further processing.

2.2.4 Microarray experiments

The Affymetrix microarray platform was used for gene expression profiling of this study. The GeneChip® HG U133 Plus 2.0, which is an oligonucleotide microarray like all Affymetrix arrays, covers probes for all transcripts of the human genome. Hence, it allows simultaneous measurement of all human genes on transcription level.

The probes on this array are single-stranded sequences of 25 nucleotides (25-mers) that are complementary to the sequence of interest. During the production of the GeneChip®, the probes are synthesized *in situ* by a photolithographic procedure. This technique was invented and patented by Affymetrix Inc. and allows the most dense packaging of features on a microarray.

In total, the HG U133 Plus 2.0 comprises 1,300,000 unique oligonucleotide features covering over 47,000 transcripts and variants, which in return represent approximately 39,000 of the best-characterized human genes.

On Affymetrix GeneChips® each transcript is represented by one or more probe sets. Each such probe set consists of 11 probe pairs, with each pair comprising $\sim 40 \times 10^7$ perfect match (PM) and mismatch (MM) probes that are organized into "probe cells" (so-called features). The perfect match probes are exactly complementary to the published sequence of the target transcript, whereas the mismatch probes exhibit a nucleotide/base exchange at position 13. Perfect match and mismatch probes of each probe pair represent different parts of the sequence of interest enabling a more robust design and analysis of the signals (Fig. 7).

During the hybridization procedure, biotin-labeled cRNA fragments of the sample to be investigated are incubated on the array and bind to their corresponding probes. The hybridization signals are generated by fluorescent-labeling of hybridized targets using phycoerythrin in a three step signal amplification process: First, a streptavidin-phycoerythrin conjugate, which binds to the biotin molecules of hybridized targets, is added. Subsequently, a biotinylated anti-streptavidin antibody is incubated to increase the effective number of biotin molecules on the target. As the last step, the streptavidin-phycoerythrin conjugate is added once more to provide binding to newly introduced biotin molecules of the anti-streptavidin antibody. Finally, the fluorescence signals of phycoerythrin are scanned. The intensity of scanned signals is representative for the abundance of target transcript within the sample.

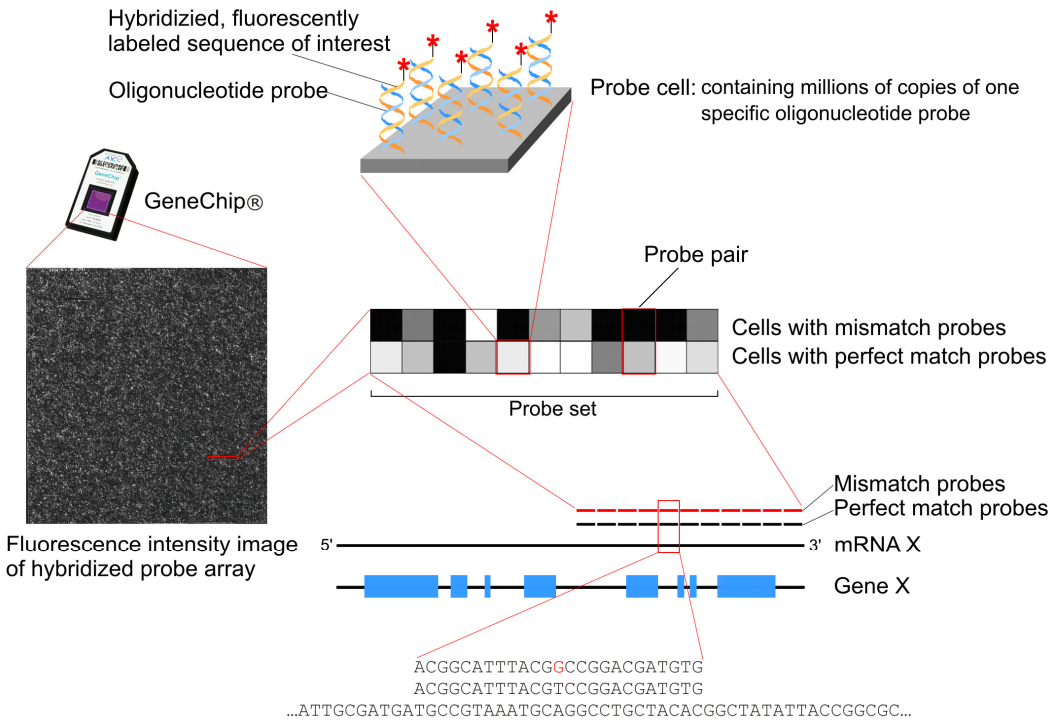


Figure 7: The Affymetrix GeneChip® technology.

In this sketch, the probe set layout of "older" GeneChip® formats, with probe pairs of one probe set being placed in rows, is depicted for simplification. However, the HG U133 Plus 2.0 array, which was used in this study, contains probe sets whose probe pairs are distributed throughout the array according to GC content of the probes. This localization strategy avoids local signal bias and improves the robustness of the platform. Nevertheless, the general principle is the same.

For microarray experiments of this study, 3 µg total RNA of each tumor sample were processed. The "One-Cycle Target Labeling" protocol ("GeneChip® Expression Analysis" technical manual, Affymetrix Inc., Santa Clara, USA) was used for sample preparation. A detailed workflow of this protocol as well as an overview of the individual products used can be found in Figure 8. In principle, this protocol comprises 3 major steps: First, DNA complementary to the whole mRNA population (cDNA) is synthesized. This step comprises an oligo(dT)-primed reverse transcription using SuperScript II for first strand synthesis, followed by second-strand cDNA synthesis using DNA Polymerase I. Secondly, after clean-up of double-stranded cDNA, *in vitro* transcription with incorporation of biotinylated pseudouridine molecules is performed using MEGASCRIPT enzyme. Finally, the obtained biotin-labeled complementary RNA (cRNA) is cleaned up and fragmented into pieces with an average size of ~100bp.

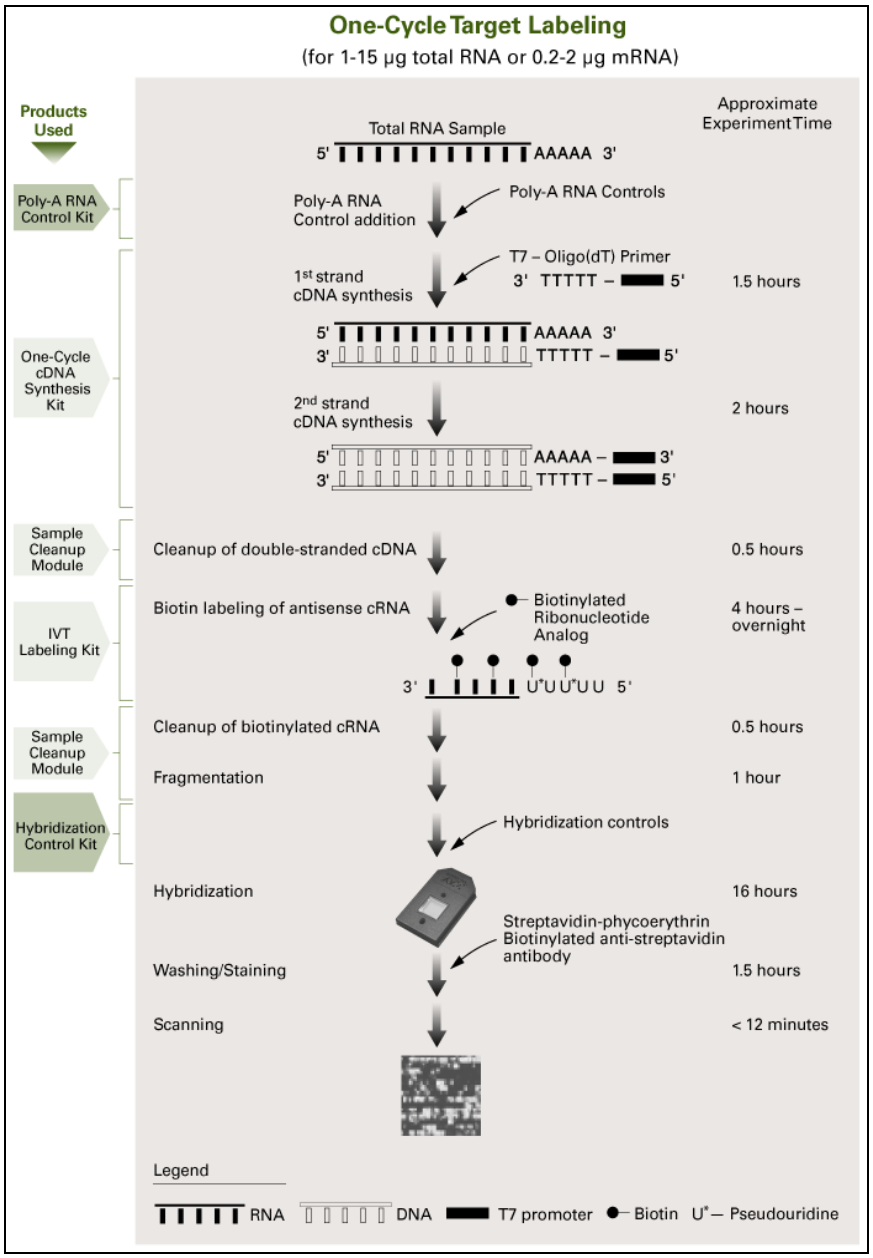


Figure 8: Workflow of "One-Cycle Target Labeling".

Adopted from "GeneChip® Expression Analysis" technical manual (Affymetrix Inc., Santa Clara, USA).

20 µg of the fragmented biotin-labeled cRNA were hybridized to the GeneChip® HG U133 Plus 2.0 followed by staining and scanning. Hybridization, staining and scanning (Affymetrix Scanner 3000) were performed in an authorized Affymetrix core facility.

2.2.5 cDNA synthesis and quantitative real-time PCR (TaqMan®-PCR)

Microarray data is inherently noisy and potentially full of artifacts, which can arise due to numerous experimental and technical reasons. Such artifacts include, for example, spurious signals caused by cross-hybridizing of single probes to sequence-similar transcripts, artifacts induced by the sample preparation technique (*e.g.* preferential enrichment of certain transcripts during RNA amplification procedures, degradation of RNA during the extraction process) and artifacts caused by the hybridization procedure itself. Additionally, large amounts of false-positives can potentially pass a statistical test, due to the high number of measurements initially introduced into the test. If, for example, all ~55,000 probe sets present on the HG U133 Plus 2.0 array are introduced into a statistical test and the *p*-value cutoff is chosen to be <0.05, which means that 5% of the probe sets pass the filter by chance, ~2,750 probe sets can be expected to pass the test as false-positives. Hence, a validation of microarray data by an independent method is indispensible.

Genes of interests extracted from the microarray data were validated by means of quantitative real-time PCR, with DNA complementary to mRNA of samples (cDNA) used as templates.

mRNA was reverse-transcribed to cDNA using SuperScript II reverse-transcriptase and oligo(dT) primers according to manufacturer's instructions. For reverse-transcription of mRNA isolated from tissues, 1 µg total RNA was employed as template. For reverse-transcription of mRNA extracted from cell, 2 µg total RNA were used. After synthesis, cDNA was diluted 1:10 with ddH_2O, aliquoted and stored at –20 °C. In parallel to each batch of cDNA synthesis, two control reactions were setup: One with ddH_2O instead of RNA (H_2O control) and a second one with ddH_2O instead of reverse transcriptase (–RT control). The H_2O control was processed to check for contamination of the cDNA synthesis reagents,

whereas the –RT control was used to check for contaminations with genomic DNA within the RNA preparation.

Due to the exponential nature of polymerase chain reaction (PCR) amplifications, the quantitation of the amplicon after each cycle can be used to calculate the initial amount of template. To do so, quantitative real-time PCRs using the TaqMan® system (Applied Biosystems, Life Technologies Corporation, Carlsbad, USA) were performed. Pre-designed "TaqMan® Gene Expression Assays", which had been optimized by the manufacturer, were purchased for all genes of interest/target genes. Table 4 (page 41) summarizes the assays used. These "Gene Expression Assays" contain target gene-specific DNA probe molecules 5'-linked to the fluorescent reporter dye 6-FAM™ (6-carboxyfluorescein) and 3'-linked to a quencher – either BHQ-1 (black hole quencher 1) or MGB (minor groove binder). Excitation of the reporter dye does not lead to fluorescence, as all emitted light is reabsorbed by the quencher in the reporter dye's close vicinity. The probes hybridize specifically to exon-exon-boundaries of the template cDNA. A *thermus aquaticus* (Taq)-polymerase-based PCR-reaction using specific PCR primers, situated up and downstream of the probe, initiates amplification. During this process, the 5' to 3' exonuclease activity of the polymerase digests the probe, separating reporter fluorophore and quencher. Thus, quencher molecules are not able to keep absorbing light emitted by the reporter. The resulting emission of light can be used to indirectly determine the amount of PCR product after each cycle by fluorescence measurement (Fig. 9). The "ABI PRISM® 7900 HT Sequence Detection System", which was used to measure the fluorescence after each cycle, subsequently plots the fluorescence intensity over the PCR cycles.

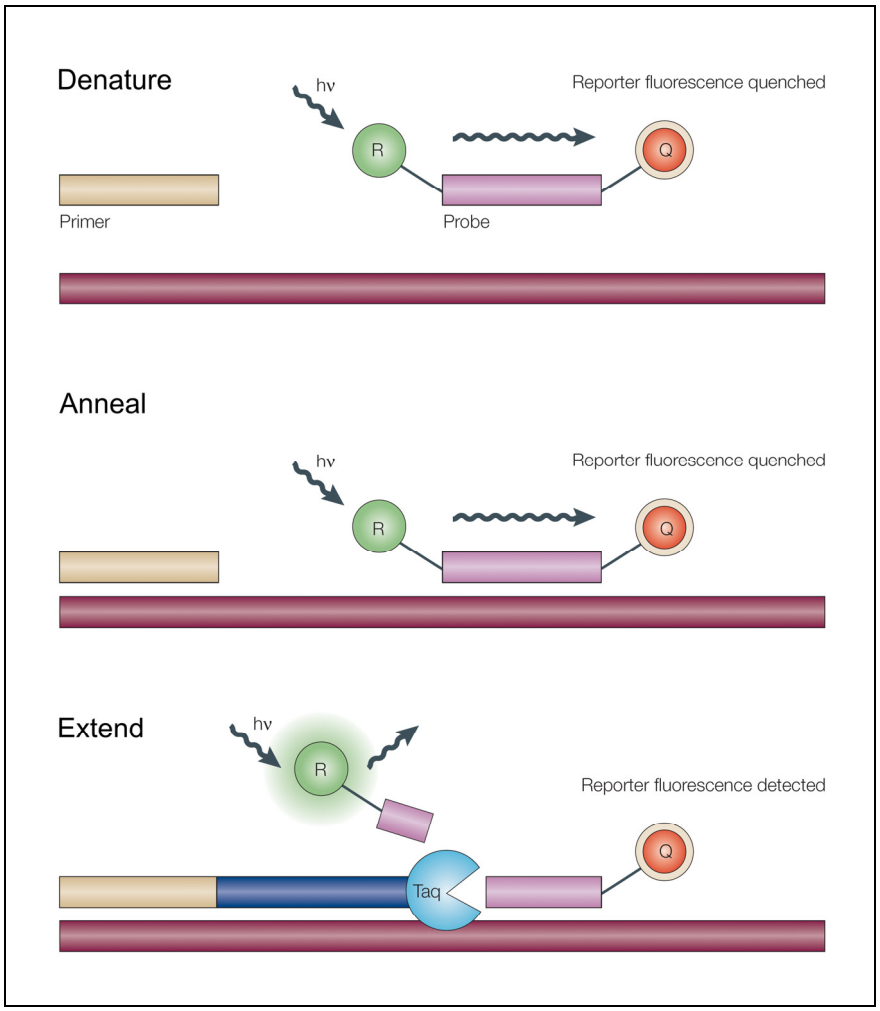


Figure 9: The principle of TaqMan®-PCR.

Figure adopted from Walter H. Koch (Koch 2004)

R – reporter fluorescent dye; Q – quencher; hυ – light quantum; Taq – *thermus aquaticus* polymerase

Additional to PCR for the target gene, each sample was also subjected to a PCR for an endogenous control ("housekeeping gene") to allow quantitation of target gene expression, later on. *Beta*-actin (*ACTB*) was chosen as the endogenous control for expression studies in gastric tumor samples. Its variation in expression across the investigated tumor samples was smaller than the one of the prominent "housekeeper" glyceraldehyde-3-phosphate dehydrogenase (*GAPDH*), a phenomenon that had been described previously elsewhere, too (Rubie *et al.* 2005). The difference in inter-tissue expression variation is illustrated in Figure 10. For

expression analyses in cell lines, *GAPDH* remained to be the endogenous control of choice.

If expression of the target gene was lower than the expression of the endogenous control, PCRs for both genes were run in duplex in the same reaction. This simultaneous detection is made possible by the availability of endogenous control expression assays comprising probes labeled with a different fluorescent reporter dye (VIC®). If the expression of the target gene was higher than the one of the endogenous control, singleplex PCRs (with amplification of both genes being performed in individual reactions) were conducted. Later on, only results from either duplex or singleplex PCRs were compared to one another.

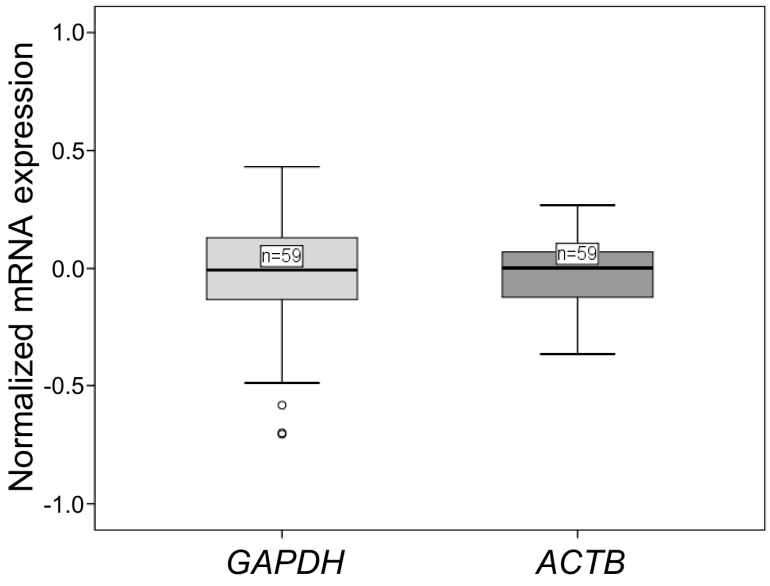


Figure 10: Inter-tissue expression variation of *GAPDH* and *ACTB* in human gastric adenocarcinomas.

Normalized (see chapter 2.2.11.1) expression values of probe sets covering the transcripts of *GAPDH* (217398_x_at, 212581_x_at, 213453_x_at) and *ACTB* (213867_x_at, 224594_x_at, 200801_x_at) were extracted from the microarray data set of all 59 samples. Means of respective expression values were calculated. Data distributions are displayed by box and whisker plots.

GAPDH – glyceraldehyde-3-phosphate dehydrogenase; *ACTB* – actin, *beta*

If available RNA amounts were "low", TaqMan®-PCRs were performed using the "RNA UltraSense™ One-Step Quantitative RT-PCR System", which is especially suitable for sensitive and reproducible detection of low-abundance RNA molecules.

Table 5 summarizes the different PCR preparation schemes used.

Table 5: Preparation of TaqMan®-PCR reactions.

conc. – concentration

"Conventional" TaqMan®-PCR:

Multiplex reaction

Component	Stock conc.	µl per reaction	Final conc.
Water		1.5	
TaqMan® Universal PCR Master Mix	2×	5.0	1×
Primer-probe-mix for target gene	20×	0.5	1×
Primer-probe-mix for endogenous control	20×	0.5	1×
Sample (cDNA)		2.5	

Singleplex reaction

Component	Stock conc.	µl per reaction	Final conc.
Water		2.0	
TaqMan® Universal PCR Master Mix	2×	5.0	1×
Primer-probe-mix for target gene or endogenous control	20×	0.5	1×
Sample (cDNA)		2.5	

"UltraSense™ One-Step"-based TaqMan®-PCR:

Multiplex reaction

Component	Stock conc.	µl per reaction	Final conc.
RNase-free water		5.0	
RNA UltraSense™ 5× Reaction Mix	5×	2.0	1×
RNA UltraSense™ Enzyme Mix		0.5	
Primer-probe-mix for target gene	20×	0.5	1×
Primer-probe-mix for endogenous control	20×	0.5	1×
ROX Reference Dye (internal reference)		0.2	
Sample (RNA)		1.3	

"Conventional" TaqMan®-PCR:

Cycle:	50.0 °C	2 min	
	95.0 °C	10 min	
	95.0 °C	15 s	←
			↑ ≥40×
	60.0 °C	1 min	→

"UltraSense™ One-Step"-based TaqMan®-PCR:

Cycle:	50.0 °C	15 min	
	95.0 °C	2 min	
	95.0 °C	15 s	←
			↑ 45×
	60.0 °C	30 s	→

Quantitation of target gene expression was performed relative to the endogenous control using the comparative C_T-method (ΔC_T-method), described in the "Guide to Performing Relative Quantitation of Gene Expression Using Real-Time Quantitative PCR" (Applied Biosystems, Life Technologies Corporation, Carlsbad, USA). In brief, the cycle, after which the signal erupts exponentially above a defined threshold, is called C_T. The C_T represents a logarithmic measure for the amount of template. To normalize the C_T of the target gene, the difference to the C_T of the endogenous control, the so-called delta C_T (ΔC_T) is calculated. The antilogarithmic value of the ΔC_T is a value for the expression of the target gene relative to the expression of the control. All TaqMan®-PCRs were performed in triplicates. Arithmetic means of C_T-values were calculated and standard error of the mean ΔC_T was calculated by taking the square root of the sum of the squared standard errors of the mean of both individual C_T values.

2.2.6 Immunohistochemistry

2.2.6.1 Single-labeling immunohistochemistry (expression/localization studies)

Frozen tissues were cut with a cryostat in 10 µm thin serial sections and mounted on SuperFrost® plus microscope slides. Sections were air dried for 2–3 hours and stored at –20 °C until use.

Frozen sections were directly transferred to –20 °C cold acetone for 5 minutes for fixation and subsequently washed with PBS (3x 5 min). After washing, sections were treated with 0.3% H_2O_2 in PBS for 5 minutes to block endogenous peroxidases. Thereafter, sections were washed again in PBS (3x 5 min) and blocked using the "Biotin Blocking System" according to the manufacturer's recommendations. Sections were treated with 0.2% Triton X-100 in PBS (5 min) for permeabilization, washed with PBS (3x 5 min) and blocked again using 3% FBS in PBS for 1 hour. After PBS washing (1x 5 min) primary antibodies were applied for 1 hour followed by further PBS washing (3x 5 min). Next, the corresponding biotinylated secondary antibody was incubated on the sections for 30 minutes at 1:500 in PBS. After subsequent washing in PBS (1x 5 min), a preformed complex of avidin and biotinylated horseradish peroxidase (ABC-complex, Vectastain Elite ABC Kit) was applied to the slides and incubated for 30 minutes. During this step, the ABC-complex binds to the biotin-molecules of secondary antibody (via avidin). Non-bound complex was eliminated by three steps of PBS washing (each 5 min). The secondary antibody-ABC-complex was finally visualized by treatment with few drops of a 3-amino-9-ethylcarbazole (AEC)-based chromogen (development solution) for approximately 10 minutes. During signal development, the AEC is oxidized by the peroxidase of ABC-complex, which results in a red coloration of antigen-positive structures and cells. Staining development was stopped by washing with distilled water. All sections were counterstained for cell nuclei with Mayer's hematoxylin (max. 1 min) plus subsequent washing with tap water. Finally, sections were mounted using fluorescence mounting medium and coverslips (aqueous mounting). All steps were performed at room temperature. Negative controls were obtained by omission of primary antibody. Primary antibodies and respective concentrations used are summarized in Table 6. Pictures were taken using the microscope Olympus BX51 and the RT KE Spot camera.

2.2.6.2 Fluorescent double-labeling immunohistochemistry (coexpression/colocalization studies)

In order to study the coexpression/colocalization of multiple proteins within a tissue sample, simultaneous detection of these proteins is necessary. Such simultaneous detection of proteins of interest is achieved by labeling with different fluorescent dyes. In this study, the fluorescent dyes Alexa Fluor® 555 and Alexa Fluor® 488, coupled to secondary antibodies, were used to detect the proteins of interest. Cell nuclei were counterstained using the fluorescent dye DAPI (4',6-Diamidin-2'-phenylindoldihydrochlorid), which selectively binds to DNA with high specificity.

Cutting of cryosections, fixation, blocking, permeabilization and associated washing steps were performed in the same manner as for single-labeling immunohistochemistry. Primary antibodies for proteins of interest were incubated on the sections for 1 hour in PBS, and sections were subsequently washed with PBS (3x 5 min). Next, the corresponding fluorescently labeled secondary antibodies and the fluorescent dye DAPI were incubated on the sections. This incubation was performed for 1 hour at concentration 1:1,000 each for antibodies and 1:2,000 for DAPI in PBS. Finally, sections were washed 3 times with PBS and mounted using fluorescence mounting medium and coverslips (aqueous mounting). All steps were performed at room temperature. Negative controls were obtained by omission of primary antibody. Primary antibodies and respective concentrations used are summarized in Table 6. Scanning of sections was done using a confocal laser scanning microscope. In the pictures, signals from Alexa Fluor® 555 are visualized in red, whereas signals from Alexa Fluor® 488 are represented in green.

Table 6: Antibodies used for immunohistochemical stainings.

Antibodies and their final concentration (in PBS) used are summarized.

IHC – immunohistochemistry; conc. – concentration; m.cl. – monoclonal; p.cl. – polyclonal; hum. – human

Protein	Single-labeling IHC	Conc.	Double-labeling IHC	Conc.
THBS4	Mouse m.cl. anti-hum. THBS4	1:100	Goat p.cl. anti-hum. THBS4	1:200
VIM	–		Mouse m.cl. anti-VIM	1:50
KRT	–		Mouse m.cl. anti-pan KRT	1:200
αSMA	–		Mouse m.cl. anti-hum. ACTA2	1:100
ProCOL1	–		Mouse m.cl. anti-hum. ProCOL1	1:100

2.2.7 *In situ* hybridization of BMP and activin membrane-bound inhibitor homolog mRNA

A well established and routinely used digoxigenin-labeled antisense probe for the BMP and activin membrane-bound inhibitor homolog (*BAMBI*) transcript was kindly provided by the group of Walter Birchmeier (Max Delbrück Center for Molecular Medicine, Berlin, Germany). It had been successfully used to detect *BAMBI* mRNA in specimens of colorectal cancer (Fritzmann *et al.* 2009). The probe had been created from the full length *BAMBI* RZPD clone IRAUp969E1147D6 (Vector pOTB7) by linearization with the restriction enzyme BamHI and subsequent amplification starting from the T7 promoter (by T7 polymerase).

In situ hybridization was performed on 5 µm thin formalin-fixed, paraffin-embedded tissue sections. All reagents used were RNase-free (DEPC-treated water was used and buffers had been autoclaved), and all incubation steps were performed at room temperature unless stated otherwise. Sections were deparaffinized by treatments with xylene for 5 minutes three times. The sections were hydrated by transfer through a descending ethanol series (100%, 95%, 85%, 70%, 50%, 30% ethanol in ddH_2O; 2 min each) and subsequent incubation in PBS for 5 minutes. Sections were refixed for 15 minutes in 4% paraformaldehyde in PBS (fresh), followed by two PBS washing-steps (5 min each). Endogenous peroxidase activity was inactivated by bathing in 6% H_2O_2 in PBS (fresh) for 15 minutes.

After another three rounds of PBS washing (each 2 min), sections were permeabilized by treatment with 10 µg/ml proteinase K in PBS (fresh) for 10 minutes, followed by incubation in 0.2% glycine in PBS (fresh) for 2 minutes. After two washings steps in PBS for 2 minutes each, sections were refixed with 4% paraformaldehyde in PBS (fresh) for 10 minutes and washed as described before. Acetylation was subsequently performed by treatment with 100mM Tris-HCl (pH 7.5) for 2 minutes followed by 10 minutes incubation with 0.25% acetic anhydride in 100mM Tris-HCl (pH 7.5; fresh).

The sections were washed twice in 2× SSC (pH 5.0; adjusted with citric acid) for 2 minutes each and dehydrated by transfer through an ascending ethanol series (30%, 50%, 70%, 85%, 95%, 100% ethanol in ddH_2O; 2 min each) and air drying (10 min–4 h). Hybridization buffer was mixed with digoxigenin-labeled probe (200 ng/ml final concentration) and ~200 µl of this mix were dispensed on each section. The sections were covered with siliconized coverslips (dip in SIGMACOTE, air dry ~20 min, dip in 100% ethanol and air dry for ~5 min) and stored upside down in a moisture chamber containing 50% formamide in 5× SSC. The probes were allowed to hybridize over night at 63 °C (the temperature depends on the GC content of the probe). The next day, coverslips were allowed to float off by bathing in 50% formamide in 5× SSC (prewarmed to 60 °C). After that, sections were washed twice in 50% formamide in 2× SSC for 10 minutes, each at 60 °C, followed by incubation in TES-buffer for 5 minutes. Two subsequent stringency washes in 1× SSC at 60 °C for 15 minutes each, were followed by one bath in 60 °C warm 0.2× SSC for 30 minutes. Sections were briefly washed in TBS (2x 5 min).

Blocking of sections, prior to signal detection, was performed by treatment with blocking solution containing Boehringer blocking reagent for 2 hours at 4 °C in a moisture chamber.

For detection of the probe, an anti-Digoxigenin antibody (Fab fragments from sheep) coupled with alkaline phosphatase was used. The antibody was blocked

using blocking solution (1:1,000) for 2 hours at 4 °C. After subsequent centrifugation at 14,000 rpm for 10 minutes at 4 °C, the supernatant was applied to the blocked sections (150–200 ml per slide) and left for incubation over night at 4 °C in a moisture chamber. The next day, sections were washed 6 times 15 minutes in 0.1% Triton X-100 in TBS and 2 times 10 minutes in NTMT. Blocking of endogenous alkaline phosphatases was achieved by incubation with 1mM levamisole in NTMT for 10 minutes. Sections were stained by application of 150–200 µl BM Purple per slide. Incubation was performed for 2 hours to 6 days at 37 °C in the dark (solution was changed every 3 days). Staining reaction was stopped in PBS. Sections were counterstained for nuclei by treatment with Nuclear Fast Red for 1 minute, followed by washing in ddH_2O for another 1 minute. Sections were dehydrated by bathing in 70% ethanol and 2 times in 100% ethanol (in ddH_2O each), followed by bathing in xylene for 3 times (1 min each). Finally, sections were mounted with entellan®.

Additional HE-counterstaining was performed on serial sections to allow improved pathological and histological evaluation of *BAMBI* positivity.

Pictures were taken using the microscope Olympus BX51 and the RT KE Spot camera.

2.2.8 Hematoxylin-Eosin-staining

Hematoxylin-Eosin (HE)-staining is one of the standard staining procedures of histology and probably the one most widely used. It enables the evaluation of tissues and their structures. In oncology, HE-stainings are usually performed to allow pathological evaluation of tissues in terms of type of malignancy, grade of differentiation, staging *etc*.

Hematoxylin has a deep blue-purple color and stains basophilic compounds, such as nucleic acids, by a complex and incompletely understood reaction. Eosin is

pink and stains eosinophilic compounds, such as proteins non-specifically. Thus, HE-staining in typical tissues results in nuclei being colored in blue, whereas the cytoplasm and extracellular matrix possess varying degrees of pink staining.

Mounted cryosections were first treated with ddH_2O for 2 minutes. Subsequently, sections were incubated in hematoxylin for 5–10 minutes (depending on amounts the solution had been used before), followed by three washing steps in tap water for 3 minutes each. Next, sections were bathed in 0.1% eosin (in ddH_2O), followed by washing in ddH_2O for 2 minutes. Sections were dehydrated by an ascending ethanol series (50%, 70%, 96%, 100%; 2 min each). After a short incubation in 2-propanol (2 min); two rounds of bathing in xylene (2 min each) were performed to completely remove any water. At the end, sections were mounted with pertex®.

2.2.9 Cell lines and cell culture

2.2.9.1 Tumor cell lines and others

Commercially available cell lines were obtained from American Type Culture Collection (ATCC), *Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH* (DSMZ) or Cell Lines Service (CLS). All cell lines were routinely cultured according to recommendations of the provider.

Non-commercially available cell lines OCUM-1, OCUM-12, OCUM-8, OCUM-2M were obtained from Osaka City University Medical School, Department of Surgical Oncology. OCUM cell lines had been established from human diffuse/poorly differentiated gastric adenocarcinomas. In culture, they predominantly grow as floating cells. Only some OCUM-2M cells are attaching.

OCUM cell lines were routinely kept in high glucose DMEM (4.5 g/l) with 10% FBS, 100 U/ml penicillin, 100 µg/ml streptomycin, 0.5 mM sodium pyruvate and 2 mM L-glutamine at 37 °C in a humidified atmosphere of 5% CO_2 in the air.

2.2.9.2 Gastric fibroblast cell lines

Cell lines of gastric cancer-associated fibroblasts (CAF) and gastric normal fibroblast (NF) were obtained from Osaka City University Medical School, Department of Surgical Oncology. They had been established from either the primary tumor site (CAF) of diffuse/poorly differentiated gastric adenocarcinoma or the non-neoplastic stomach (NF) of the same patient, as matched pairs. In brief, the tissue material had been excised under aseptic conditions and minced with forceps and scissors. The tissue pieces were cultured in high glucose DMEM with 10% FBS, 100 U/ml, 100 µg/ml streptomycin, 0.5 mM sodium pyruvate and 2 mM L-glutamine at 37 °C in a humidified atmosphere of 5% CO_2 in the air in a 100 mm culture dish. The fibroblasts grew as monolayers and were collected and transferred to another culture dish every 5–7 days. The fibroblast monocultures were verified by immunostaining of vimentin and cytokeratin with monoclonal antibodies (Dako, Glostrup, DEN). The procedure is described elsewhere (Yashiro *et al.* 1996).

Fibroblast cell lines were routinely cultured in high glucose DMEM with 10% FBS, 100 U/ml penicillin, 100 µg/ml streptomycin, 0.5 mM sodium pyruvate and 2 mM L-glutamine at 37 °C in a humidified atmosphere of 5% CO_2 in the air.

2.2.10 Indirect coculture experiments

The growth of distinct cell types in a combined *in vitro* culture is referred to as "coculture". Coculture experiments are generally used to examine the effects of one specific cell type (cell line) to another one (or *vice versa*) or the interactions between them. They serve as an *in vitro* modeling approach for *in vivo* cell interactions (Lackie 2007). Examples of interactions to be monitored include effects of certain secreted proteins of one cell type (*e.g.* growth factors) to another one and effects of cell-to-cell contact mechanisms, such as cell adhesion.

Coculture experiments can be conducted in a direct or indirect fashion, depending on the interaction type to be studied and the feasibleness of the experiment. The direct approach, which features combined culturing of the cell types to be investigated in the same culture dish, is "coculture" by definition. This straight forward approach is somewhat "easy" to conduct, but bears some disadvantages. For example, the separation of the different cell lines at the end of the experiment, which is necessary for certain cell type-specific downstream applications, can be tedious and difficult to achieve. In indirect formats, the interaction of cells is enabled by secreted material without combined culturing of cell types. One example for such an approach is the culture of different cell lines in different chambers separated by a certain barrier-like semipermeable membrane. Such membranes allow the exchange of media and its components from one chamber to the other, but not the passing through of cells. Depending on the size of pores, different fractions of proteins can pass the membrane or are retained. A second example of indirect coculture makes use of the conditioned medium of the cell line whose effects on other cell lines should be tested (= the donor cell line). The conditioned medium is the cell culture medium in which the donor cells have grown for a certain period of time. It is therefore depleted of some compounds, which have been partially used up by the donor cell line, but is enriched by donor-specific material, such as secreted proteins (*e.g.* growth factors, enzymes). Later on, this donor cell-

conditioned medium is transferred to the cell line whose reaction to the donor is the objective of the experiment (= the recipient/acceptor cell line). Due to the experimental nature, indirect coculture approaches can only monitor the interactions and effects caused by secreted material. Effects caused by direct cell-to-cell contacts can not be assessed.

For this thesis, indirect coculture experiments were conducted using the conditioned medium technique. The objective of these experiments was to study the reaction of different fibroblasts (recipients) to different tumor cells (donors). For preparation of tumor cell-conditioned medium, cells were suspended in medium at a concentration of $6–7\times10^4$ cells/ml. After 48 hours of culture, the medium was centrifuged for 5 minutes at 840*g* to remove cells and cell debris and the supernatant was collected. All conditioned media were aliquoted and stored at –80 °C until use. For indirect coculture experiments, $2\times10^4–1\times10^5$ fibroblasts were seeded in wells of 6-well plates. After 24 hours of initial culture, the fibroblast medium was replaced by the tumor cell-conditioned medium (2ml/well). At 48 hours after incubation with conditioned medium, fibroblasts were harvested, lysed in RLT buffer (RNeasy Micro Kit) and stored at –80 °C until use in downstream applications. A simplified scheme of the procedure can be found in Figure 11. The non-conditioned medium (original/fresh medium) of respective tumor cell line was used as control medium. All experiments were performed in triplicates and the control experiments were run on the same plate, respectively. Similar approaches and protocols are described elsewhere (Guo *et al.* 2008, Nakagawa *et al.* 2004).

Downstream total RNA extraction from fibroblasts was performed using the QIAShredder columns and the RNeasy Micro Kit. Expression of genes of interest was examined by means of quantitative real-time PCR ("UltraSense™ One-Step"-based TaqMan®-PCR).

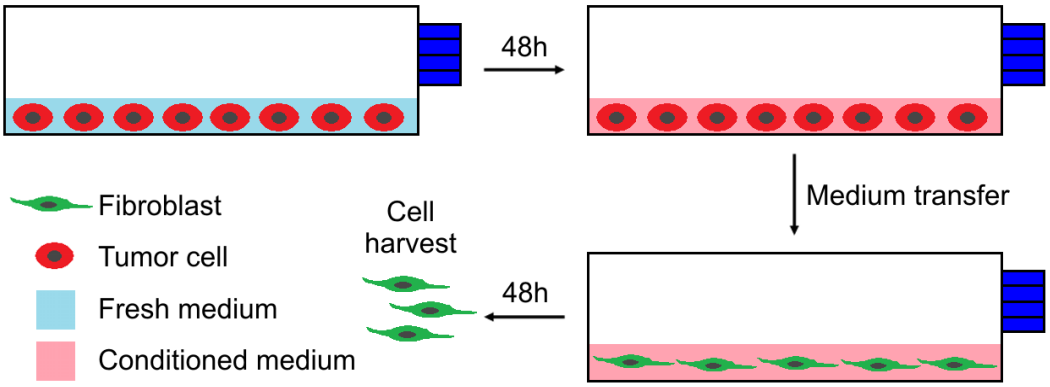


Figure 11: Principle of indirect coculture experiments performed in this study.

Medium that had been conditioned by tumor cells for 48 hours was transferred to fibroblast cultures. After 48 hours of incubation with tumor cell-conditioned medium, fibroblast cells were harvested, lysed and stored at –80 °C until downstream extraction of total RNA.

h – hour/s

2.2.11 Bioinformatic and statistical microarray data analysis

All microarray data analyses were performed in GeneSpring GX 10.0.2 software, unless stated otherwise.

2.2.11.1 Data preprocessing

Preprocessing of Affymetrix microarrays involves three key steps: background correction/adjustment, normalization and probe summarization.

For preprocessing, original CEL-data files were imported into GeneSpring GX and the GC-RMA (GC-Robust Multi-array Average) algorithm (Wu *et al.* 2004) was applied. It includes the mentioned three major steps of background correction, normalization and probe summarization. The GC-RMA method is an improved form of RMA (Bolstad *et al.* 2003, Irizarry *et al.* 2003a, Irizarry *et al.* 2003b) regarding background correction. In contrast to RMA, the background correction

implemented within the GC-RMA algorithm takes different tendencies of probes to encounter non-specific binding (based on GC content) into consideration, thus avoiding underestimation of background noise. These sequence-specific probe affinities are completely ignored by the background correction done by the original RMA method. Normalization and probe summarization of RMA and GC-RMA are exactly the same. For probe summarization, signal intensities of mismatch probes are ignored and the summarized expression value of each probe set is created based on the perfect match probes, only.

The expression measures outputted by the GC-RMA algorithm are on log2 scale. Finally, these measures were baseline transformed to the median of all samples. During this step, the median value of each probe sets across all arrays is computed (on log scale) and subtracted from each individual value of this probe set.

2.2.11.2 Quality control

In order to verify the quality of the microarray data, a correlation analysis across GeneChips® was performed. The correlation coefficient of each GeneChip® to all other GeneChips® was calculated and coefficients were plotted against each other as a heatmap (correlation plot). This type of analysis allows estimating on a visual basis whether outlier chips are present within the data set or whether all chips are comparable to one another.

2.2.11.3 Comparison of independent groups – Identification of significantly differentially expressed gene

Different clinical parameters (*e.g.* histological type, recurrence of disease) were used to form patient groups. Significantly differentially expressed genes between groups were identified by Welch-test, a modified unpaired *t*-test, which does not assume variances of samples of each group to be equal.

In a typical microarray study several thousands of genes are simultaneously introduced to a statistical test. During the test, each gene is considered independently from one another, and thus the test is performed on each gene separately. The incidence of false-positives (genes passing the test, but possess no difference between groups in reality) is proportional to the number of tests performed and the critical significance level chosen (*p*-value cutoff). The *p*-value is the probability that the gene passes the test due to chance alone. A *p*-value of 0.05 signifies a 5% probability that the gene passes the test by chance. If, for example, 10,000 genes are tested, 5% or 500 genes might be called significant just by chance. This is why it is important to correct the *p*-value of each gene, when performing a statistical test on a group of genes. Multiple testing correction algorithms, which are used for this purpose, correct the individual *p*-value for each gene to keep the overall error rate (or false-positive rate) to less than or equal to the user-desired *p*-value cutoff or error rate.

In this study, multiple testing correction techniques of Benjamini and Hochberg False Discovery Rate (Benjamini and Hochberg 1995) and Bonferroni Family Wise Error Rate (Bonferroni 1935, Bonferroni 1936) were used.

Statistical significance was accepted at corrected $p<0.05$. If multiple testing corrections resulted in no significant features passing the test, statistical significance was accepted at non-corrected $p<0.001$.

The fold change (FC) of expression between two groups was calculated as the fold difference between group means.

2.2.11.4 Gene Ontology analysis

The results of high-throughput experimental techniques like microarrays are lists/groups of interesting genes, *e.g.* lists of differentially expressed genes, which require further biological interpretation and evaluation. One possibility to accom-

plish this task is to use the gene-specific functional annotations provided by the Gene Ontology (GO) system.

For this study, GO analysis was performed using GOSSIP, a freely available software package that tests whether a molecular function, biological process or cellular location, described in the Gene Ontology system (the so-called GO terms), is significantly associated with a group of interesting genes when compared to a reference group (Bluthgen *et al.* 2005). As a result, lists of statistically enriched GO terms are outputted. In order to avoid misleading results, GOSSIP implements multiple testing corrections when determining statistical significance. Thus, GOSSIP represents a powerful and reliable tool to identify and examine the biological relevance of gene groups of interest.

2.2.11.5 Clustering analyses

Clustering is the assignment of observations/objects of one data set into subsets, the so-called clusters, such that those within the same cluster are more closely related/similar to one another than to the objects in the other clusters. Central to all the goals of cluster analysis is the intention to identify similarity (or dissimilarity) between the individual objects investigated.

In microarray studies the primary goal is to find "co-behaving" subsets of genes or samples. Clustering of genes/probe sets is performed based on the expression profile of each individual probe set across all microarrays, and probe sets that "behave" similar (*e.g.* up or downregulated in the same arrays) are clustered together. For clustering of samples/microarrays, the expression values of a certain set of genes/probe sets is used, and individual microarrays are clustered according to their expression profile. As the result, microarrays exhibiting similar expression profiles are assigned to the same cluster.

All clusterings were done by means of hierarchical clustering algorithms. The result of such hierarchical cluster analyses is a tree-like structure, the so-called dendrogram, which displays the hierarchy of the formed clusters. “Euclidian distance” and “complete linkage” were used as distance metric and linkage algorithm for all clustering analyses performed (clustering of microarrays/samples and of probe sets/genes).

2.2.12 Statistical evaluation of quantitative real-time PCR data

All statistical evaluations of quantitative real-time PCR data were performed using SPSS 16.0 software.

2.2.12.1 Comparison of independent groups

In order to evaluate whether relative levels of mRNA are significantly differentially distributed between groups (patients/tumor samples), non-parametric statistical tests were performed. Depending on whether two or more groups were compared to one another, Mann-Whitney-U-tests or Kruskal-Wallis-H-tests were conducted. These non-parametric tests for comparison of independent data sets make no assumption about the distribution of data (*e.g.* normality), unlike the parametric *t*-test. Thus, they represent alternatives to the independent group *t*-test, when the assumption of normality or equality of variance is not met.

The *p*-value of asymptotic significance was chosen as significance estimate, with two-tailed significance being estimated in Mann-Whitney-U-tests.

2.2.12.2 Estimation of accuracy and performance of diagnostic tests

The diagnostic performance of a test or the accuracy of a test to discriminate certain cases from other ones can be evaluated using receiver operating characteristic (ROC) curve analysis (Metz 1978, Zweig and Campbell 1993). The result of any particular test system in two populations will rarely separate the two populations perfectly (to 100%). Instead, the distribution of the test results will overlap to a certain extent. Thus, for every possible threshold point (classifier boundary) value that can be selected to discriminate between the two populations, there will be some cases correctly and some cases incorrectly assigned to their respective group. For binary classification systems (two-class prediction problem), in which the test outcomes are labeled either as positive or negative, four types of outcome are possible (Fig. 12): Correctly classified positives (true-positives) or negatives (true-negatives) and incorrectly classified positives (false-positives) or negatives (false-negatives). Parameters giving information about the ratio of true-positives and false-positives at a certain threshold point are the true-positive rate (TPR) and the false-positive rate (FPR). The TPR determines the ratio of positive cases correctly classified among all positive cases available during the test, whereas the FPR defines how many incorrect positive results occur among all negative cases available during the test.

In a ROC curve the TPR (sensitivity) is plotted as a function of the FPR (1-specificity) for different cutoff points (Fig. 13). Each point on the ROC plot represents a sensitivity-specificity pair corresponding to a particular decision threshold. Hence, the ROC curve is a visual index of the accuracy of a test. A test with perfect discrimination (no overlap in the two distributions) has a ROC curve passing through the upper left corner with 100% sensitivity (no false-negatives) and 100% specificity (no false-positives). Thus, the closer the curve passes by this upper left corner, the higher the overall accuracy of the test (Zweig and Campbell 1993). The calculated area under the ROC curve gives an estimate about how close the

actual curve is to the one of an ideal test, with a value of 1 being equal to the ideal curve.

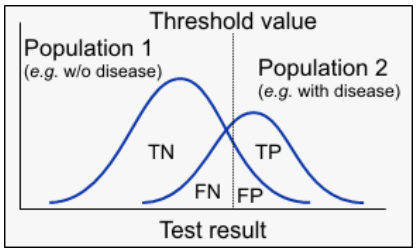


Figure 12: Schematic distribution of test results of binary classification systems in two populations.

True-positives (TP), false-positives (FP), true-negatives (TN) and false-negatives (FN) are depicted according to selected threshold value.

w/o – without; *e.g.* – for example

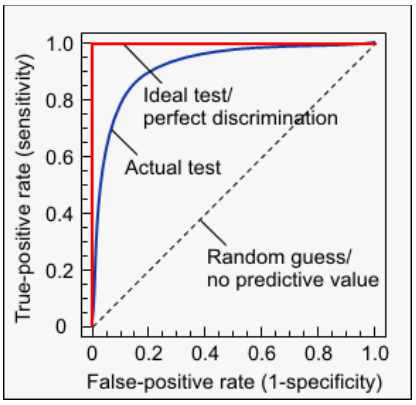


Figure 13: Schematic overview of possible ROC curves.

ROC curves from tests with different accuracies are depicted.

In this work, receiver operating characteristic was used to evaluate the performance of a gene's relative mRNA expression level to potentially serve as diagnostic tests distinguishing patients with recurrence or formation of postoperative metastases from relapse-free patients. According to ROC curve the most "appropriate" cutoff value of expressions was selected with aiming for high sensitivity and low specificity of the test, respectively.

2.2.12.3 Survival analysis

Survival analyses were performed using the Kaplan-Meier method (Kaplan and Meier 1958). This procedure is a common way to summarize survival data via the estimation of survival probabilities at certain time points. The Kaplan-Meier survival curve plots the proportion of survival/probability of survival as a function of

time, with each death being represented by a downward step in the curve. Thus, the curve shows, for each time point, the proportion of subjects that survive at least this length of time.

An important advantage of the Kaplan-Meier method is that it can take "censored" data (losses of samples before the final outcome is observed) into account. In clinical applications such censored data can occur when a patient withdraws from a study, for instance. In the curve, small vertical ticks indicate such “losses”.

The time until postoperative relapse/progression, the so-called disease-free survival (DFS) or the time until formation of postoperative metastases, the so-called metastasis-free survival (MFS) were chosen for survival analyses.

Differences in survival times between groups were compared using the non-parametric logrank-test. Significance in survival difference was accepted at $p<0.05$.

2.2.12.4 Multivariate analysis

Multivariate analyses are techniques to analyze more than one variable at a time.

Binary logistic regression is the method of choice if your dependent variable (so-called criterion variable) is binary (dichotomous) and you wish to explore the relative influence of continuous and/or categorical independent variables (so-called covariates) on your dependent variable, and to assess interaction effects between the independent variables (Spicer 2004).

In this thesis, binary logistic regression was used to estimate the independence of the information offered by genes of interest (on transcriptional level) from the information provided by other clinical parameters.

3 Results

3.1 Microarray data quality control

General quality of the microarray data set was assessed by means of correlation. The correlation coefficient of each individual GeneChip® to every other GeneChip® was calculated and coefficients were plotted against each other as a heatmap (Fig. 14). Correlation coefficients ranged between 0.85 and 0.98 with a mean of 0.92 (median=0.917). No outlier microarray/s with low correlation to other arrays could be recognized.

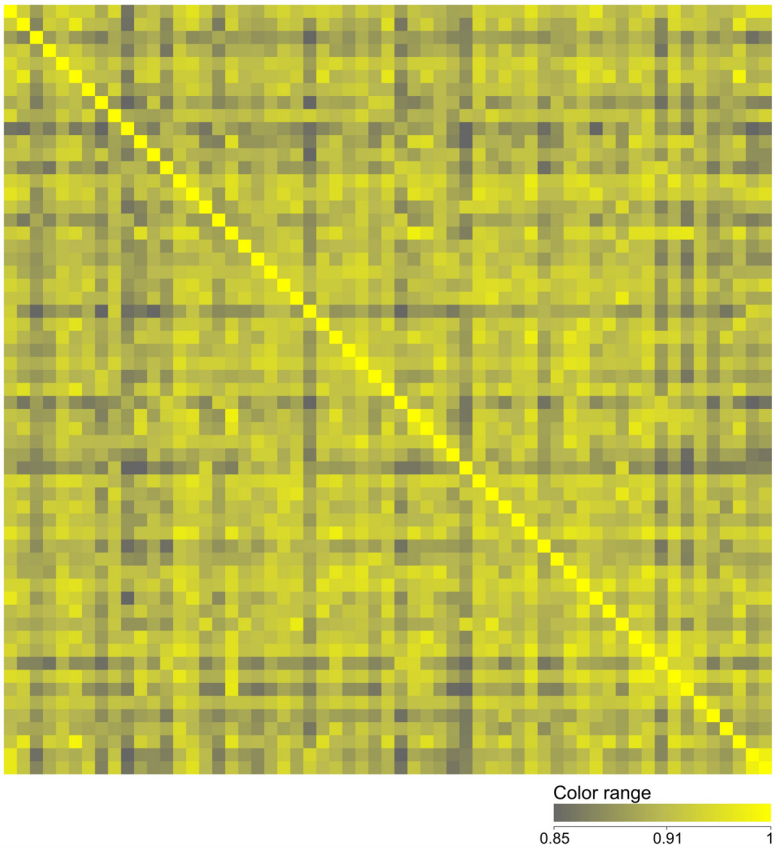


Figure 14: Correlation plot of all microarrays.

The correlation coefficient of each GeneChip® to every other GeneChip® was calculated and coefficients were plotted against each other as a heatmap.

3.2 Identification of relevant subgroups of gastric adenocarcinoma

Unsupervised hierarchical clustering of microarrays was used to discover relevant trends and subgroups of tumor samples/patients whose members share certain similarities in gene expression profile and thus act as biological subgroups.

First, the expression values of all genes (54,675 probe sets) were used to cluster the tumor samples. This whole gene expression profile-based approach revealed that two prominent clusters of samples exist within the data set. Examination for correlation with clinicopathological parameters elucidated that these two subgroups are generally determined by the histological type of adenocarcinoma (Fig. 15). Most tumors of the intestinal type belonged to one of the clusters (87.9%), whereas the majority of diffuse-type tumors had been assigned to the other cluster (69.2%). In a second approach, only probes sets possessing raw signal intensity values (linear values after summarization) within the range of the 50^{th} and 100^{th} percentile in at least 10 microarrays, namely 29,505 probe sets, were used to cluster the samples. The application of this expression-based pre-filtered list of genes resulted in a dendrogram showing the same two histology-defined tumor clusters (Fig. 15). However, the separation of the two subgroups was even slightly stronger with 93.9% of intestinal and 73.1% of diffuse samples being assigned to their respective cluster.

Correlation of the observed tumor sample clusters with any other clinicopathological parameter, including T-stage, N-stage, M-stage, age, gender, prognostic parameters (recurrence of disease, formation of postoperative metastases, …) *etc*., could not be identified.

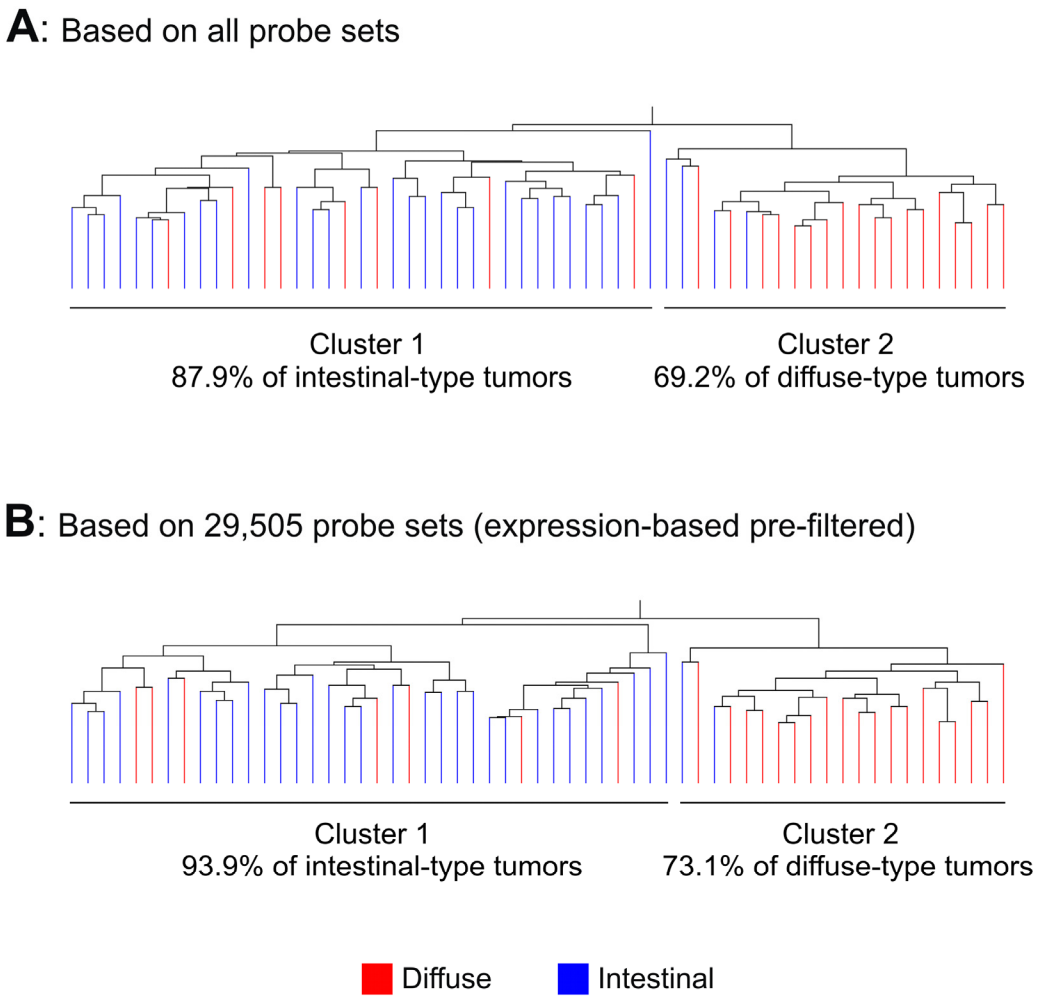


Figure 15: Unsupervised hierarchical clustering of human gastric adenocarcinoma samples with respect to histological type.

A: Dendrogram (hierarchical tree) of samples as obtained by application of expression values of all probe sets (54,675) for clustering. B: Dendrogram (hierarchical tree) of samples obtained when using expression values of an expression-based pre-filtered list of 29,505 probe sets (linear signal intensity values after summarization within the range of the 50th and 100th percentile in at least 10 arrays) for clustering. "Euclidian distance" and "complete linkage" were used as distance metric and linkage algorithm for both clusterings, respectively.

3.3 Thrombospondin 4 – the most prominent member of gene signatures for histological type of gastric adenocarcinoma

3.3.1 Establishment of global gene expression profiles of diffuse and intestinal-type gastric adenocarcinoma

Unsupervised hierarchical clustering had elucidated that diffuse and intestinal-type adenocarcinomas exhibit strong differences in gene expression profile. To filter out which individual genes are differentially expressed between these two histological groups, statistical significance testing comparing each gene's average expression in the diffuse group to its average in the intestinal group (Welch-test) was performed. Only samples featuring tumor contents of >75% and a clear pathological evaluation were used to form the histological groups (19 diffuse, 24 intestinal). In order to reduce numbers of false-positives passing the test, multiple testing correction was applied during the test. Only genes/probe sets yielding corrected *p*-values <0.05 and fold changes ≥2 were regarded as differentially expressed.

Depending on the multiple testing method used, different numbers of significantly differentially expressed probe sets were identified. Application of Bonferroni Family Wise Error Rate (FWER), the most conservative and stringent multiple testing technique available, yielded the smallest amount of probe sets, namely 322. These 322 probe sets covered 207 unique annotated transcripts and 44 unique non-annotated ones (NetAffx, September 2009). The more moderate technique of Benjamini and Hochberg False Discovery Rate (FDR) identified 2,071 probe sets representing 1,280 unique annotated transcripts and 253 unique non-annotated ones (NetAffx, September 2009) to be significantly differentially expressed. In both test systems, the vast majority (*e.g.* ~73 % for FDR) of identified differentially expressed genes were upregulated in diffuse-type gastric tumors, whereas

upregulation in the intestinal type applied to a smaller amount of genes (*e.g.* ~27% for FDR), only (Fig. 16). Annotated lists of the 50 genes possessing the most significant upregulation in either histological type can be found in the appendix (Tab. 17 page 165 ff. and Tab. 18 page 170 ff.).

Application of the obtained gene signatures for hierarchical clustering of tumor samples resulted in dendrograms comprising two major sample clusters representing the two histological types, consequentially. However, usage of the 2,071 probe set list gained by Benjamini and Hochberg FDR resulted in 95.3% (41 of 43) of samples being grouped to the respective cluster, whereas the more stringently selected 322 probe set list of Bonferroni FWER produced a dendrogram in which 97.7% (42 of 43) were grouped "properly" (Fig. 16).

The trade-off and drawback of very stringent multiple testing corrections, such as the Bonferroni FWER, are that big numbers of false-negatives may occur within the test system. Thus, a lot of genes that are in fact differentially expressed between the groups investigated may not pass the test and "become lost". For this reason, the genes identified as being differentially expressed by Benjamini and Hochberg FDR were used for biological interpretation of data.

In order to find out which biological connection is present among the individual genes overexpressed in either histological type, Gene Ontology (GO) analysis was performed. This investigation uncovered that genes overexpressed in intestinal-type adenocarcinomas are predominantly associated with proliferation and growth connected processes, such as cell cycle and mitosis (Tab. 7). In contrast, most of the genes upregulated in diffuse-type adenocarcinomas encode for proteins of the extracellular matrix and/or for proteins that play important roles in adhesion or developmental processes (Tab. 8). No overlap of significantly enriched GO terms could be identified for the genes strongly expressed in either histological group.

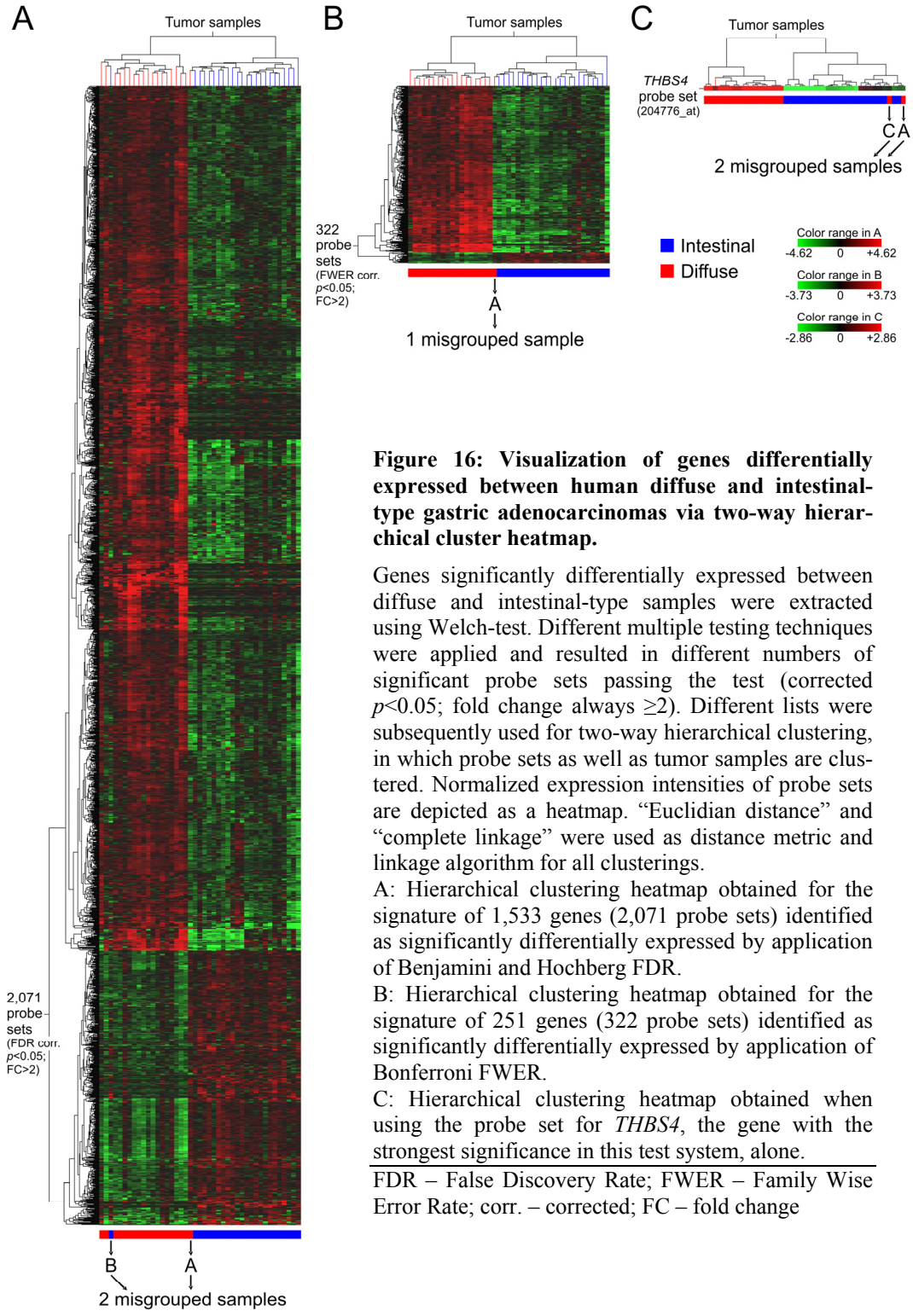


Figure 16: Visualization of genes differentially expressed between human diffuse and intestinal-type gastric adenocarcinomas via two-way hierarchical cluster heatmap.

Genes significantly differentially expressed between diffuse and intestinal-type samples were extracted using Welch-test. Different multiple testing techniques were applied and resulted in different numbers of significant probe sets passing the test (corrected $p<0.05$; fold change always ≥ 2). Different lists were subsequently used for two-way hierarchical clustering, in which probe sets as well as tumor samples are clustered. Normalized expression intensities of probe sets are depicted as a heatmap. "Euclidian distance" and "complete linkage" were used as distance metric and linkage algorithm for all clusterings.

A: Hierarchical clustering heatmap obtained for the signature of 1,533 genes (2,071 probe sets) identified as significantly differentially expressed by application of Benjamini and Hochberg FDR.

B: Hierarchical clustering heatmap obtained for the signature of 251 genes (322 probe sets) identified as significantly differentially expressed by application of Bonferroni FWER.

C: Hierarchical clustering heatmap obtained when using the probe set for *THBS4*, the gene with the strongest significance in this test system, alone.

FDR – False Discovery Rate; FWER – Family Wise Error Rate; corr. – corrected; FC – fold change

Table 7: Significantly enriched GO terms identified for genes overexpressed in human intestinal-type gastric adenocarcinomas.

Gene Ontology analysis was performed using GOSSIP (Bluthgen *et al*. 2005).

GO – Gene Ontology; ID – identifier; FDR – False Discovery Rate; NDF – Number of Degrees of Freedom; # – number of probe sets; ref. – reference

GO ID	GO term description	*p*-value	FDR corr. *p*-value	NDF corr. *p*-value	# in test list	# in ref. list
GO:0022403	cell cycle phase	1.72E-60	5.57E-58	5.57E-58	88	565
GO:0022402	cell cycle process	4.38E-60	6.56E-58	1.31E-57	95	711
GO:0000279	M phase	1.00E-59	1.11E-57	3.32E-57	79	419
GO:0000278	mitotic cell cycle	3.08E-55	2.58E-53	1.03E-52	81	525
GO:0007049	cell cycle	4.94E-53	3.51E-51	1.76E-50	103	1067
GO:0000087	M phase of mitotic cell cycle	8.80E-53	4.88E-51	2.93E-50	66	309
GO:0005694	chromosome	6.04E-36	3.38E-34	2.36E-33	73	789
GO:0044427	chromosomal part	2.61E-34	1.30E-32	1.04E-31	69	735
GO:0051726	regulation of cell cycle	5.81E-27	2.81E-25	2.53E-24	52	523
GO:0000075	cell cycle checkpoint	2.22E-26	9.35E-25	9.35E-24	27	81
GO:0005819	spindle	1.24E-24	5.09E-23	5.60E-22	23	54
GO:0007067	mitosis	5.83E-23	2.14E-21	2.57E-20	27	115
GO:0043228	non-membrane-bound organelle	6.44E-23	2.14E-21	2.84E-20	124	3458
GO:0043232	intracellular non-membrane-bound organelle	6.44E-23	2.14E-21	2.84E-20	124	3458
GO:0044446	intracellular organelle part	2.79E-22	8.44E-21	1.27E-19	150	4827
GO:0044422	organelle part	4.16E-22	1.18E-20	1.89E-19	150	4847
GO:0007059	chromosome segregation	2.75E-21	7.55E-20	1.28E-18	20	49
GO:0000819	sister chromatid segregation	4.62E-21	1.22E-19	2.19E-18	18	33
GO:0015630	microtubule cytoskeleton	7.21E-21	1.81E-19	3.44E-18	49	641
GO:0007346	regulation of mitotic cell cycle	2.87E-20	6.88E-19	1.38E-17	23	92
GO:0006323	DNA packaging	4.94E-20	1.14E-18	2.39E-17	35	310
GO:0006259	DNA metabolic process	2.44E-19	5.07E-18	1.11E-16	56	923
GO:0051276	chromosome organization and biogenesis	6.70E-19	1.43E-17	3.28E-16	52	814
GO:0006996	organelle organization and biogenesis	3.61E-16	7.57E-15	1.82E-13	86	2335
GO:0051325	interphase	8.20E-16	1.63E-14	4.07E-13	24	175
GO:0031570	DNA integrity checkpoint	1.08E-15	2.15E-14	5.58E-13	17	64
GO:0000785	chromatin	7.22E-15	1.37E-13	3.71E-12	29	309
GO:0044430	cytoskeletal part	1.07E-14	1.94E-13	5.42E-12	53	1079
GO:0051329	interphase of mitotic cell cycle	4.34E-14	7.94E-13	2.30E-11	22	171
GO:0005874	microtubule	4.61E-14	8.06E-13	2.42E-11	14	44
GO:0031497	chromatin assembly	4.56E-13	7.82E-12	2.42E-10	26	289
GO:0007017	microtubule-based process	1.04E-12	1.81E-11	5.78E-10	29	382
GO:0006333	chromatin assembly or disassembly	1.52E-12	2.54E-11	8.39E-10	26	306
GO:0000226	microtubule cytoskeleton organization and biogenesis	4.67E-12	7.45E-11	2.53E-09	19	154
GO:0051327	M phase of meiotic cell cycle	5.38E-12	8.29E-11	2.90E-09	17	116
GO:0007093	mitotic cell cycle checkpoint	5.46E-12	8.29E-11	2.96E-09	12	40
GO:0051321	meiotic cell cycle	6.08E-12	8.80E-11	3.25E-09	17	117
GO:0051301	cell division	6.80E-12	9.37E-11	3.56E-09	15	83
GO:0005856	cytoskeleton	3.91E-11	5.50E-10	2.15E-08	57	1520
GO:0043231	intracellular membrane-bound organelle	8.84E-11	1.26E-09	5.03E-08	244	12726

Table 8: Significantly enriched GO terms identified for genes overexpressed in human diffuse-type gastric adenocarcinomas.

Gene Ontology analysis was performed using GOSSIP (Bluthgen *et al*. 2005).

GO – Gene Ontology; ID – identifier; FDR – False Discovery Rate; NDF – Number of Degrees of Freedom; # – number of probe sets; ref. – reference

GO ID	GO term description	*p*-value	FDR corr. *p*-value	NDF corr. *p*-value	# in test list	# in ref. list
GO:0031012	extracellular matrix	8.43E-63	3.51E-60	3.51E-60	137	484
GO:0005576	extracellular region	5.99E-54	8.95E-52	2.69E-51	189	1158
GO:0044421	extracellular region part	5.99E-54	8.95E-52	2.69E-51	189	1158
GO:0048856	anatomical structure development	3.40E-45	4.27E-43	1.71E-42	347	3798
GO:0032501	multicellular organismal process	4.38E-44	4.58E-42	2.29E-41	475	6245
GO:0007275	multicellular organismal development	1.88E-42	1.69E-40	1.01E-39	328	3584
GO:0032502	developmental process	4.19E-39	3.22E-37	2.25E-36	438	5787
GO:0022610	biological adhesion	7.64E-38	5.14E-36	4.11E-35	188	1529
GO:0048731	system development	4.72E-37	2.84E-35	2.55E-34	300	3333
GO:0005578	proteinaceous extracellular matrix	1.55E-34	8.48E-33	8.48E-32	62	164
GO:0044420	extracellular matrix part	2.71E-34	1.42E-32	1.56E-31	62	166
GO:0009653	anatomical structure morphogenesis	8.98E-31	4.24E-29	5.09E-28	189	1768
GO:0048513	organ development	2.87E-28	1.42E-26	1.85E-25	229	2511
GO:0007155	cell adhesion	1.86E-21	9.20E-20	1.29E-18	100	776
GO:0005515	protein binding	5.64E-19	2.59E-17	3.88E-16	349	5332
GO:0003012	muscle system process	6.78E-18	3.03E-16	4.84E-15	53	281
GO:0005581	collagen	1.15E-16	5.00E-15	8.51E-14	24	46
GO:0030247	polysaccharide binding	2.71E-16	1.09E-14	1.97E-13	39	165
GO:0043062	extracellular structure organization and biogenesis	3.90E-16	1.51E-14	2.88E-13	42	196
GO:0001568	blood vessel development	6.44E-16	2.29E-14	4.58E-13	54	329
GO:0007399	nervous system development	1.79E-15	6.59E-14	1.38E-12	104	1031
GO:0001871	pattern binding	5.15E-15	1.83E-13	4.02E-12	39	183
GO:0030198	extracellular matrix organization and biogenesis	7.25E-15	2.43E-13	5.59E-12	20	34
GO:0001944	vasculature development	1.10E-14	3.53E-13	8.48E-12	57	392
GO:0022008	neurogenesis	3.19E-14	1.00E-12	2.51E-11	73	617
GO:0030246	carbohydrate binding	1.14E-13	3.45E-12	8.96E-11	59	443
GO:0000902	cell morphogenesis	1.25E-13	3.72E-12	1.00E-10	73	636
GO:0032989	cellular structure morphogenesis	2.62E-13	7.43E-12	2.08E-10	74	661
GO:0048699	generation of neurons	2.93E-13	8.11E-12	2.35E-10	68	577
GO:0030029	actin filament-based process	4.65E-13	1.23E-11	3.68E-10	60	473
GO:0009887	organ morphogenesis	1.15E-12	2.97E-11	9.22E-10	72	654
GO:0008092	cytoskeletal protein binding	3.29E-12	8.16E-11	2.61E-09	77	745
GO:0016477	cell migration	3.78E-12	9.69E-11	3.20E-09	59	486
GO:0035295	tube development	5.58E-12	1.34E-10	4.57E-09	47	332
GO:0031589	cell-substrate adhesion	6.97E-12	1.61E-10	5.64E-09	38	226
GO:0005539	glycosaminoglycan binding	9.89E-12	2.28E-10	8.21E-09	31	153
GO:0035239	tube morphogenesis	1.11E-11	2.43E-10	8.99E-09	38	230
GO:0022604	regulation of cell morphogenesis	2.07E-11	4.38E-10	1.66E-08	24	91
GO:0007154	cell communication	2.32E-11	4.93E-10	1.92E-08	392	6990
GO:0030182	neuron differentiation	3.35E-11	7.04E-10	2.82E-08	53	432

3.3.2 Thrombospondin 4 – the most potent marker for histological type of gastric adenocarcinoma in this data set

3.3.2.1 *THBS4* in the microarray data

Among the identified genes possessing significant mRNA overexpression in diffuse-type adenocarcinomas, thrombospondin 4 (*THBS4*) exhibited the highest fold change and lowest *p*-value. It was 40.8 fold upregulated in these tumors compared to intestinal-type ones and held a corrected (Benjamini and Hochberg FDR) *p*-value of 1.65E-7. Hence, this gene was chosen for further analyses.

Hierarchical clustering of tumor samples performed based on the *THBS4* probe set alone yielded the same sample grouping like when using the gene signature obtained by Benjamini and Hochberg FDR (Fig. 16).

3.3.2.2 Validation of *THBS4* microarray data via quantitative real-time PCR

Eleven samples of each histological group were randomly chosen to validate the *THBS4* microarray data by means of quantitative real-time PCR. In parallel, *THBS4* mRNA expression was examined in the two "misgrouped" samples (refer to Fig. 16) as well to correct for potential hybridization artifacts *etc.* on these microarrays. This analysis clearly validated the microarray data and confirmed the strong significance in differentially *THBS4* mRNA expression between the two histological types ($p<0.0001$, Mann-Whitney-U-test). Furthermore, it pinpointed that *THBS4* mRNA is principally absent from the examined intestinal-type gastric adenocarcinoma specimens, whereas varying amounts are present within the diffuse-type population. Interestingly, *THBS4* mRNA abundance in the "misgrouped" samples was observed to rank right in between the amounts encountered

in diffuse and intestinal type (Fig. 17). Weak statistical significance could be noticed when compared to the group of diffuse or intestinal type (p=0.03, Mann-Whitney-U-test).

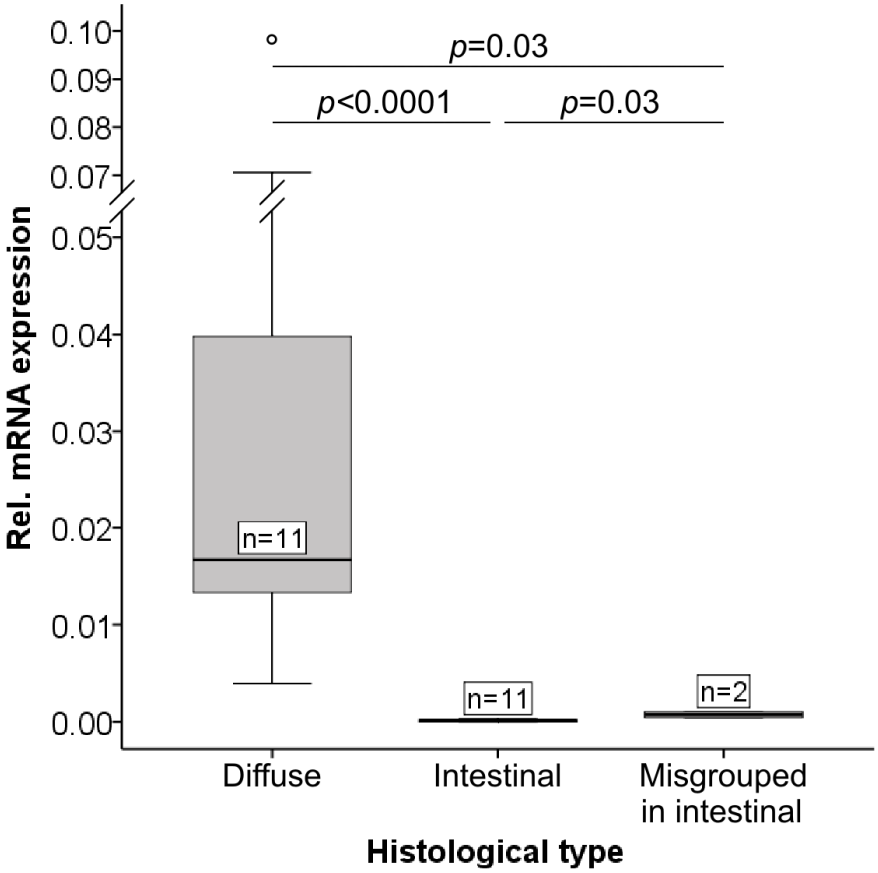


Figure 17: *THBS4* mRNA expression in human diffuse and intestinal-type gastric adenocarcinomas.

The mRNA abundance of *THBS4* was examined in 11 randomly chosen samples of each histological type by means of quantitative real-time PCR. Quantitation was done relative to the transcript of *β*-actin. Significance in differential expression between groups was calculated using Mann-Whitney-U-test and *p*-value of two-tailed asymptotic significance was chosen as significance estimate. Expression value distribution within groups is displayed by box and whisker plot.

Rel. – relative; n – number

3.3.2.3 THBS4 expression in diffuse and intestinal-type gastric adenocarcinomas

To investigate whether the differences on mRNA level are reflected on protein level and to identify the cellular localization of THBS4 protein expression within diffuse-type adenocarcinomas, immunohistochemistry experiments were performed.

All examined diffuse-type tumors showed specific positivity for THBS4 protein (Fig. 18 & 19), whereas no notable specific staining could be observed in the investigated intestinal tumor specimens (Fig. 20).

The main localization of THBS4 within diffuse-type tumors could be identified to be the extracellular matrix of the tumor stroma. All investigated specimens showed THBS4 positivity of extracellular fibrillar structures surrounding the tumor cells (Fig. 18). In general, THBS4 positivity was particularly strong within regions of high tumor cell density. For example, examined diffuse tumor case 1 (Fig. 18), which represents a mucinous diffuse-type adenocarcinoma, showed explicitly intense positivity within tumor cell nests surrounded by extracellular mucus. Another example is diffuse tumor case 4 (Fig. 19), which likewise exhibited stronger positivity within regions of tumor cell accumulation than in regions of low to moderate tumor cell density. At sites of tumor infiltration into adjacent "healthy" epithelium, very strong THBS4 expression could be observed (case 2B, Fig. 18) as well, whereas more moderate expression was present within the tumor mass itself (case 2A, Fig. 18).

In the stroma of few diffuse-type tumors, additional intracellular positivity for THBS4 could be detected (Fig. 19). Cells possessing this cytosolic THBS4 expression were rather small in size and had a fusiform or spindle-like shape with sometimes extended cell processes; all features which denote a potential fibroblast phenotype. No clear cytosolic THBS4 positivity could be encountered in the rather big and roundly shaped tumor cells.

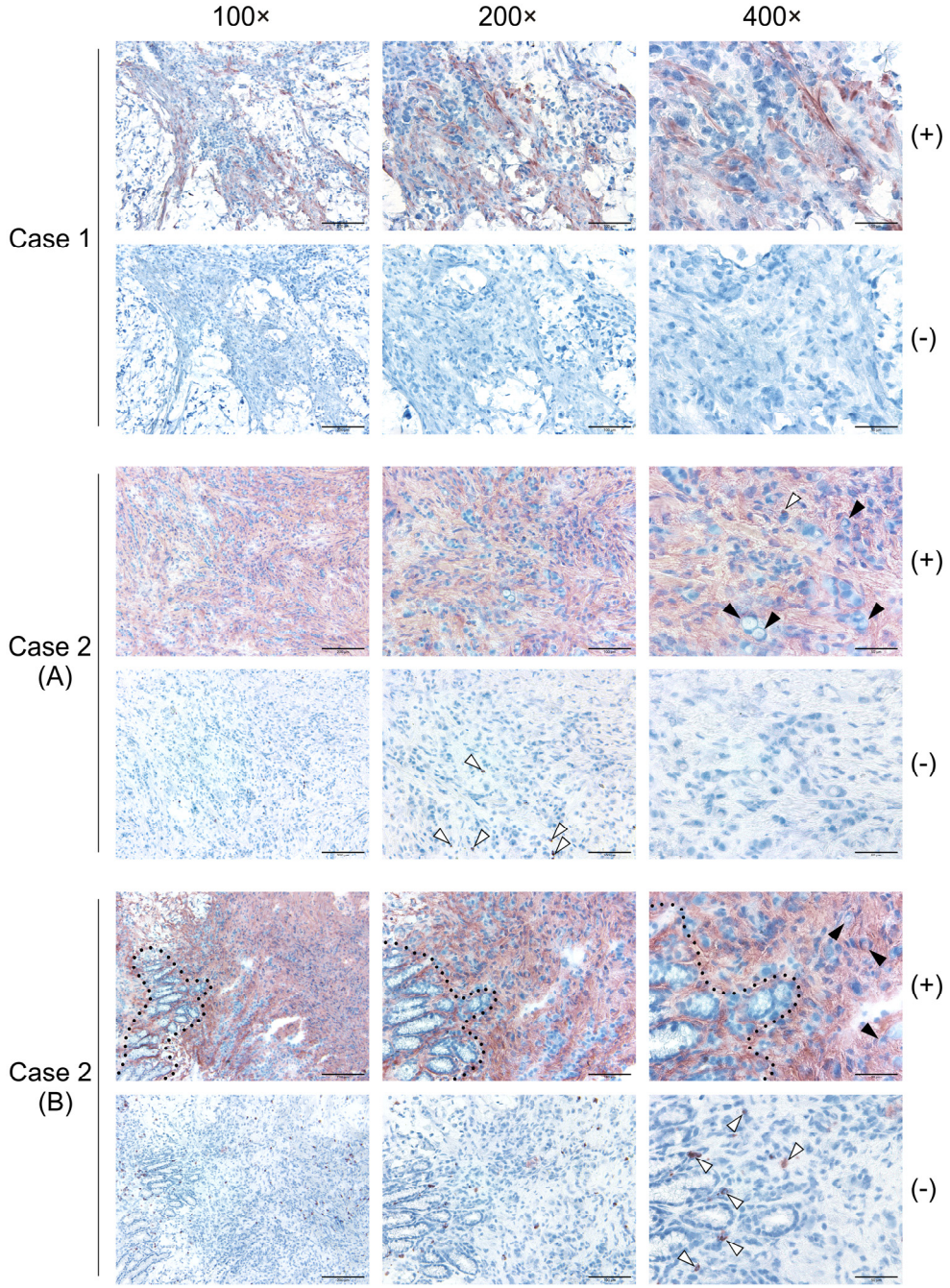


Figure 18: THBS4 expression in human diffuse-type gastric adenocarcinomas (extracellular localization).

Immunohistochemical detection of THBS4 (red) was performed on 10 µm thin cryosections. Cell nuclei were counterstained using hematoxylin (blue). Negative controls (-) were obtained by omission of primary antibody. Representative sections possessing THBS4 positivity in extracellular fibrillar structures of the tumor stroma are depicted. Case 1 = a mucinous-type tumor (mucus is visible as white areas); case 2 = a signet ring cell carcinoma (remnants of healthy epithelium are marked by dotted lines; examples of signet ring cells are depicted by black arrowheads; white arrowheads mark unspecific staining of immune cells, which was occasionally encountered)

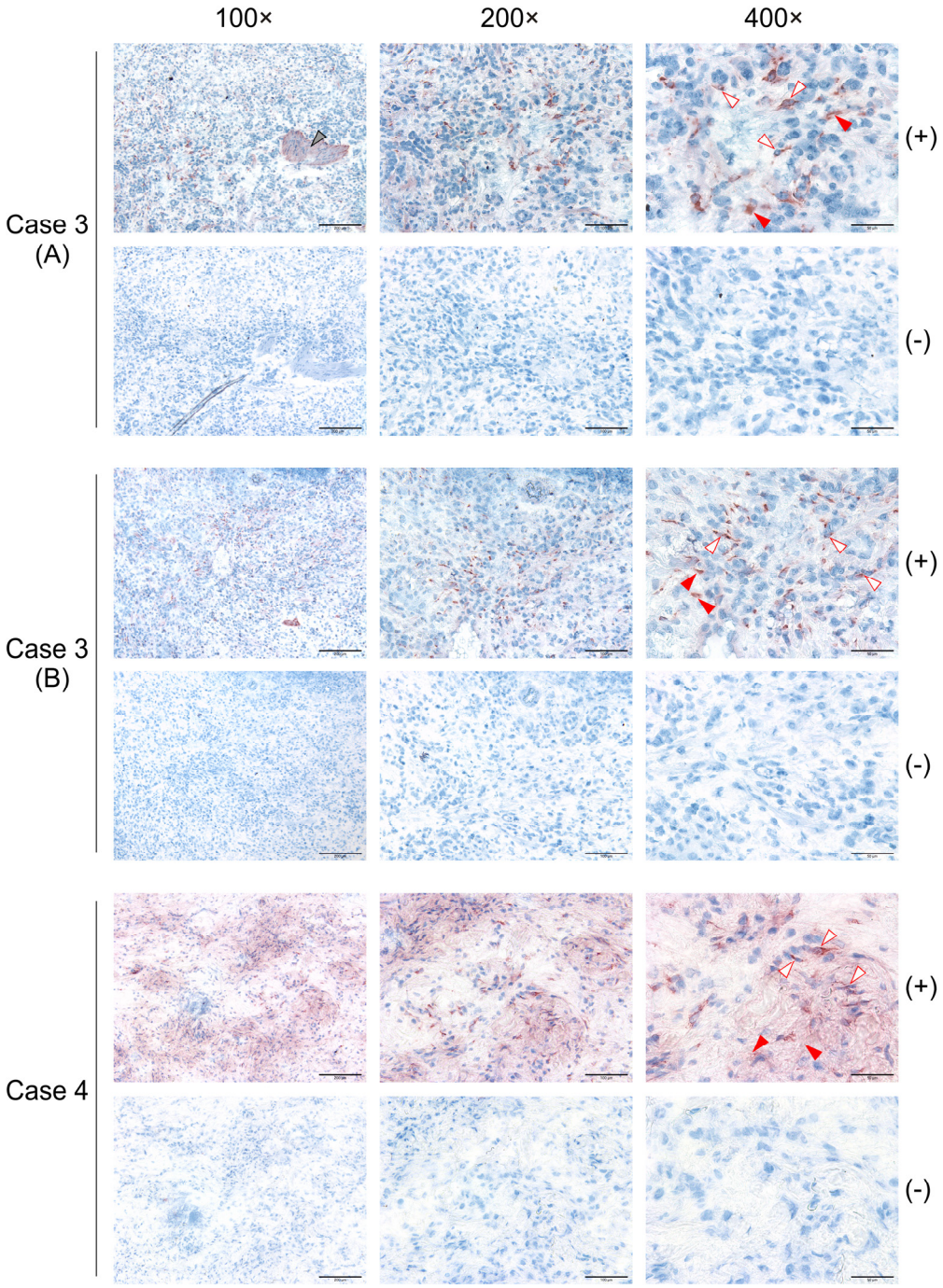


Figure 19: THBS4 expression in human diffuse-type gastric adenocarcinomas, *continued* (cellular localization).

Immunohistochemical detection of THBS4 (red) was performed on 10 µm thin cryosections. Cell nuclei were counterstained using hematoxylin (blue). Negative controls (-) were obtained by omission of primary antibody. Representative sections possessing cytosolic positivity (red open arrowheads) additional to extracellular positivity (red arrowheads) for THBS4 are depicted. The grey arrowhead in case 3 (A) marks remnants of smooth muscle, which also exhibit THBS4 expression (refer to Fig. 22).

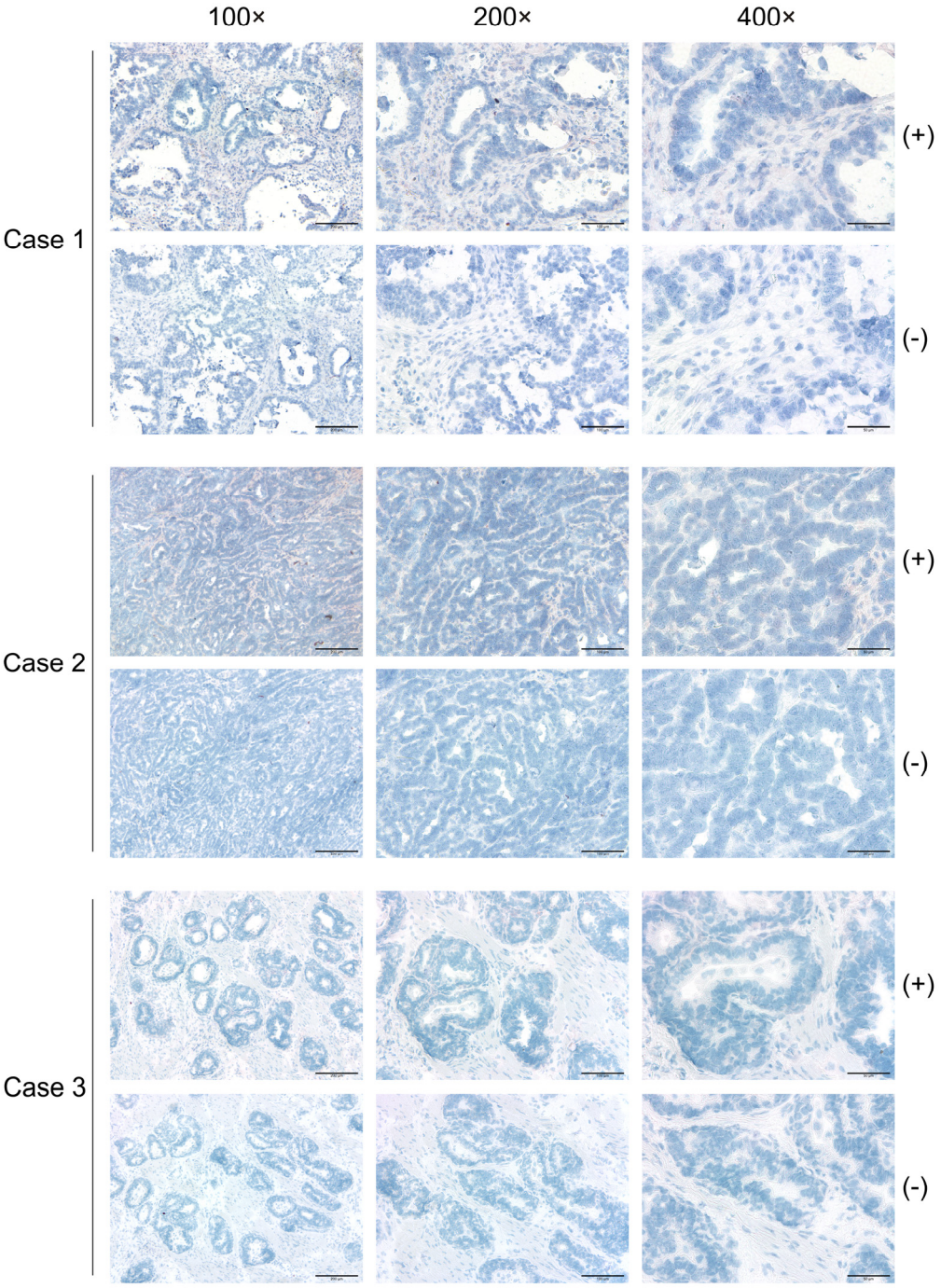


Figure 20: THBS4 expression in human intestinal-type gastric adenocarcinomas.

Immunohistochemical detection of THBS4 (red) was performed on 10 µm thin cryosections. Cell nuclei were counterstained using hematoxylin (blue). Negative controls (-) were obtained by omission of primary antibody. No specific THBS4-positivity could be observed in all investigated samples. Representative sections are shown.

3.3.2.4 THBS4 expression in “normal” non-neoplastic gastric tissue

For expression studies in “normal” gastric tissue, matched (from the same patient) specimens of non-neoplastic gastric tissue, which had been resected adjacent to the tumor (at the tumor margins), were used. These experiments revealed that epithelium and stroma (connective tissue) of “normal” gastric mucosa do not express THBS4. Likewise, no THBS4 expression could be detected in the submucosal stroma. The only aspects of the gastric wall, which could be proven to display THBS4 expression, are the muscularis mucosae and muscularis propria as well as vessel walls. THBS4 expression in these smooth muscle layers was noticed to be restricted to the extracellular connective tissue in between muscle cells. The smooth muscle cells themselves did not show any clear positivity for the protein (Fig. 21 & 22).

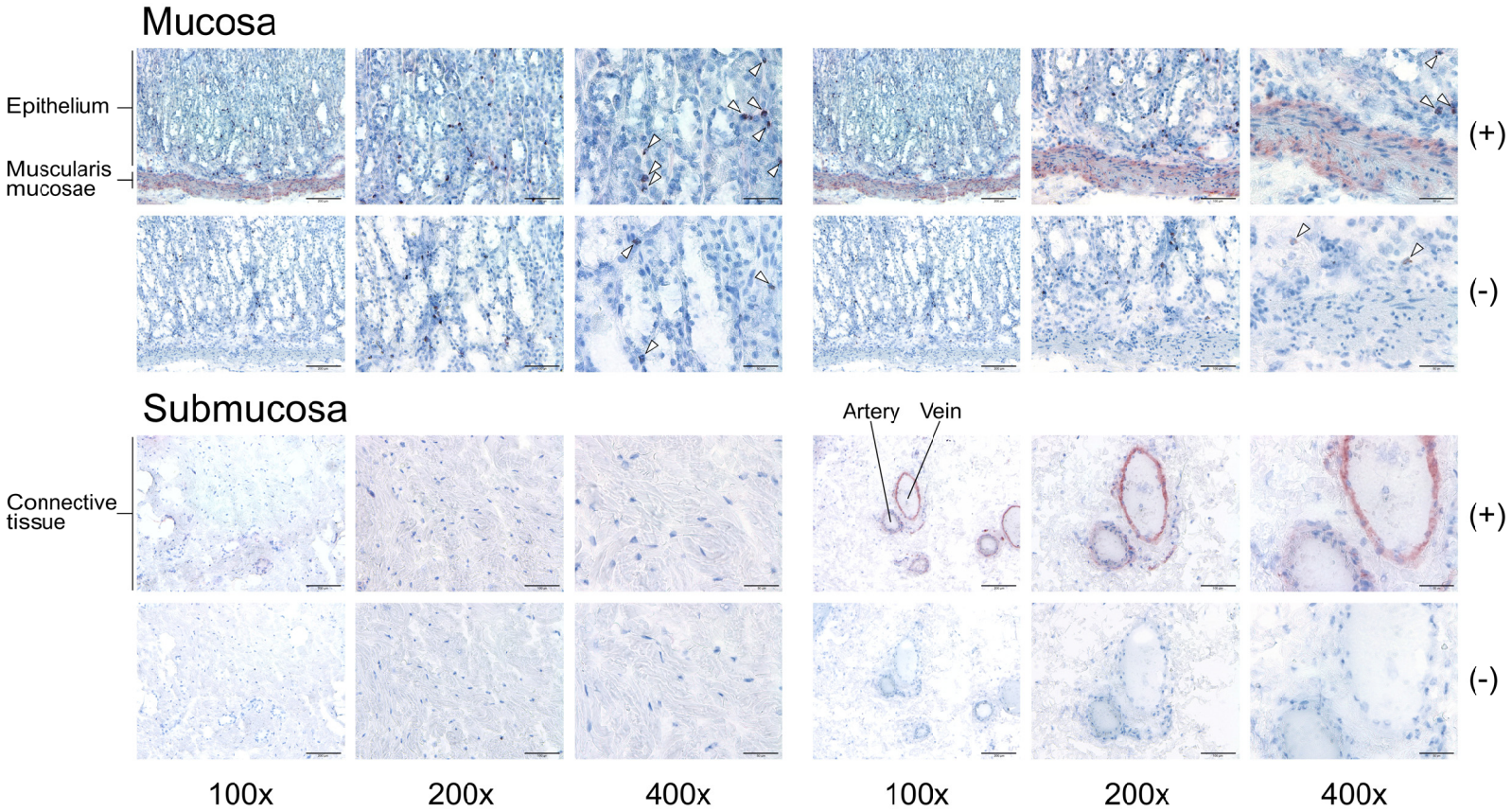


Figure 21: THBS4 expression in human non-neoplastic gastric tissue (*figure continues on next page*).

Immunohistochemical detection of THBS4 (red) was performed on 10 µm thin cryosections covering the full thickness of the gastric wall (except the serosa). Cell nuclei were counterstained using hematoxylin (blue). Negative controls (-) were obtained by omission of primary antibody. THBS4 expression can only be observed in muscularis mucosae (extracellularly in between smooth muscle cells) and in vessel walls. No THBS4 expression is visible in epithelium and connective tissue of mucosa and submucosa. White arrowheads mark unspecific staining of immune cells, such as macrophages, which was occasionally encountered.

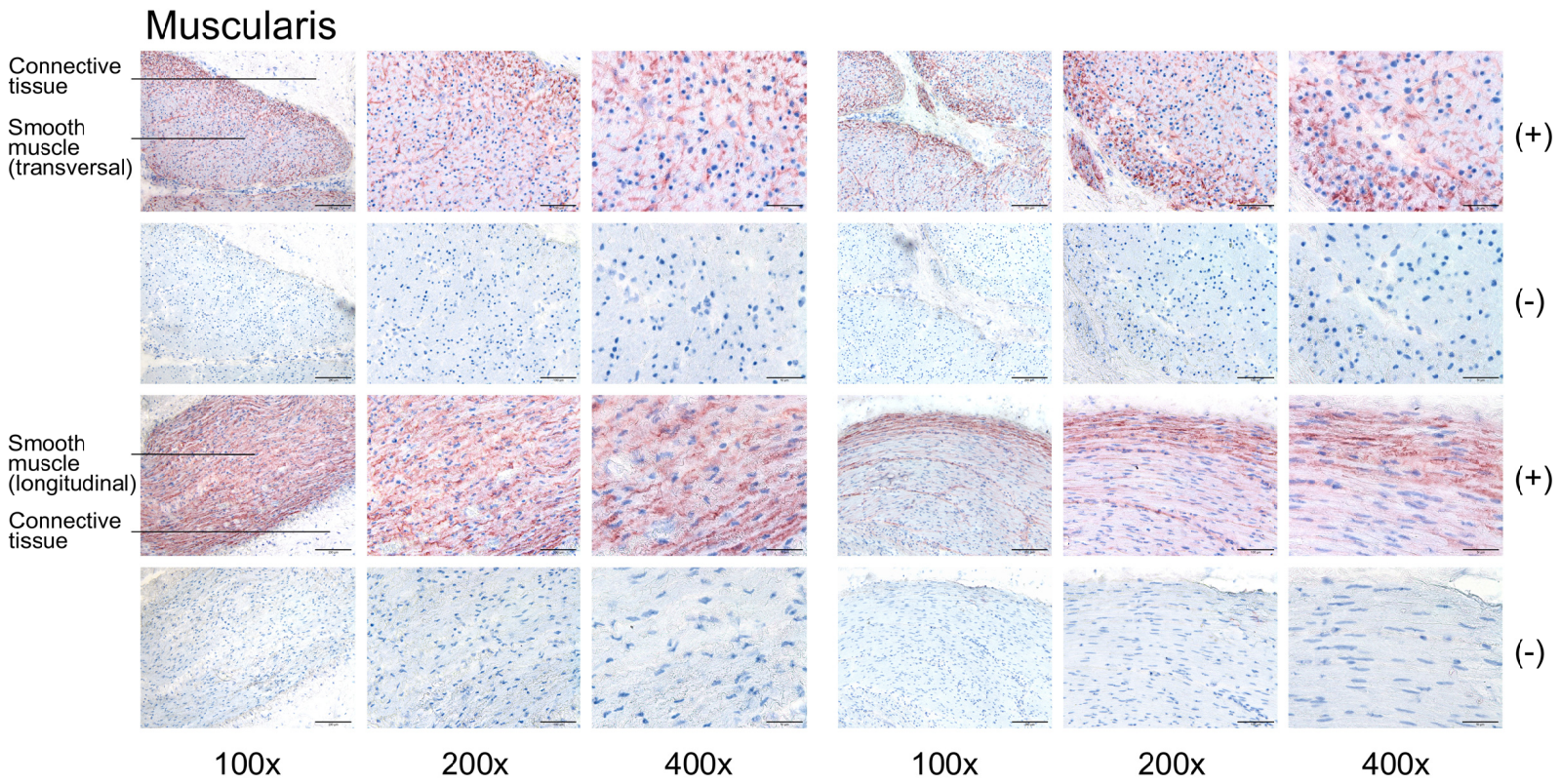


Figure 22: THBS4 expression in human non-neoplastic gastric tissue, *continued.*

Immunohistochemical detection of THBS4 (red) was performed on 10 µm thin cryosections covering the full thickness of the gastric wall (except the serosa). Cell nuclei were counterstained using hematoxylin (blue). Negative controls (-) were obtained by omission of primary antibody. THBS4 expression can be observed in all layers of muscularis propria (extracellularly in between smooth muscle cells).

3.3.2.5 Determination of the cellular origin of extracellular THBS4 in diffuse-type gastric adenocarcinomas

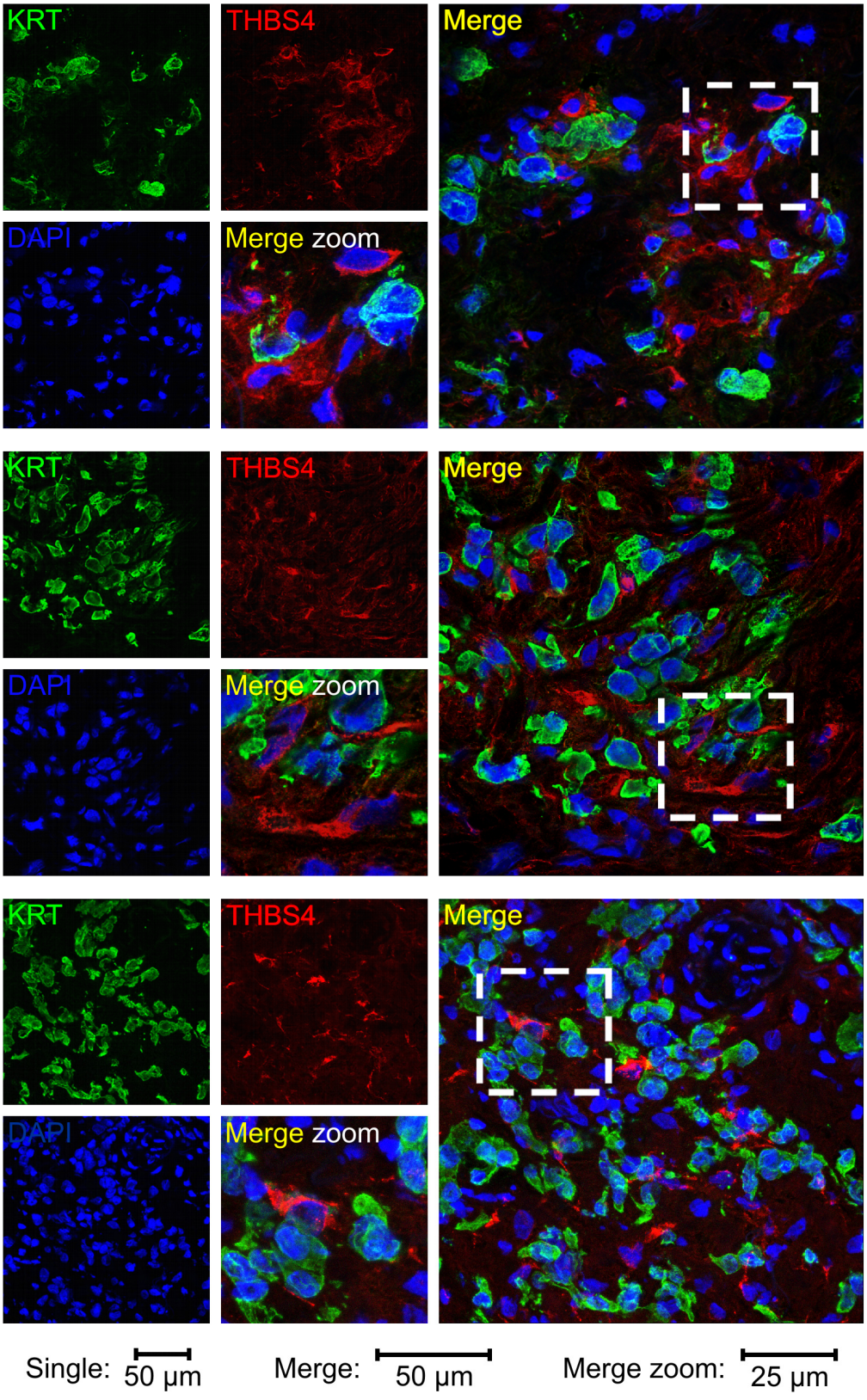
Given that the cells featuring cytosolic THBS4 positivity are presumably the cells that express and secrete THBS4, and that therefore are the origin of extracellular THBS4 accumulation in diffuse-type gastric tumors, it is indispensable to identify their cellular entity. According to morphology, they resembled fibroblasts. However, a clear identification of the entity is not possible on the basis of such morphological estimations alone. Therefore, colocalization studies using markers for different cell entities were conducted.

Cytokeratins, keratin-containing intermediate filaments of the cytoskeleton of epithelial cells, served as markers for cells of epithelial origin, in this regard the tumor cells. For denoting fibroblasts, a panel of different proteins was used, because "one" specific and reliable molecular fibroblast marker does not exist or has not been identified so far. There are several well-established indicators of the fibroblast phenotype, but none of them are both exclusive to fibroblasts and present in all fibroblasts (Kalluri and Zeisberg 2006). The panel of fibroblast markers used in this study included vimentin, *α*-smooth muscle actin and procollagen 1. Vimentin, a member of the intermediate filament protein family of the cytoskeleton of mesenchymal cell, was used to signify mesenchymal cells, of which fibroblast represent a subpopulation of. *Alpha*-smooth muscle actin, which inherently is a constituent of the contractile apparatus of smooth muscle cells, served as an indicator for myofibroblasts. Myofibroblasts are activated fibroblasts with features that are more typical of smooth muscle differentiation, such as possessing bundles of contractile microfilaments (consisting of actins and myosins) and maintaining gap junctions. Due to this ambivalent smooth muscle-like phenotype, they had been termed myofibroblasts by G. Gabbiani in 1971 (Gabbiani *et al*. 1971). Myofibroblasts are found in normal tissue as well as in a wide variety of pathological processes. In the context of malignant transformation, they represent a subpopulation of cancer-associated fibroblasts. Procollagen 1, which is an intracellular pre-

cursor of collagen I, was used as a general comprehensive fibroblast marker in this study.

For colocalization studies, immunohistochemistry experiments with fluorescent double-labeling of THBS4 and either one of the described marker proteins were performed. Specimens used were the diffuse tumor samples, in which cytosolic THBS4 positivity had been observed in single-labeling immunohistochemical experiments.

The co-labeling with cytokeratins directly proved that THBS4 expression in diffuse-type tumors is an event taking place in the near vicinity of tumor cells. Extracellular and intracellular/cytosolic THBS4 expression could be encountered close to the tumor cells in all investigated samples, with regions of high tumor cell density showing exceeding levels of expression (Fig. 23). Cells positive for THBS4 predominantly exhibited a fibroblast-typical spindle-like shape with rather small flattened nuclei (compared to tumor cells) and long extensions, as observed before. None of these THBS4-positive cells showed positivity for cytokeratins in any of the analyzed tumors, thus confirming that the tumor cells themselves are not the origin of extracellular THBS4 (Fig. 23 & 24).



for figure legend see next page

Figure 23: Coexpression of THBS4 and cytokeratin, a marker for carcinoma cells, in human diffuse-type gastric adenocarcinomas.

Simultaneous fluorescent immunohistochemical detection of THBS4 in red and pan cytokeratin (KRT4, 5, 6, 8 10, 13 and 18) in green was performed on 10 µm thin cryosections. Cell nuclei were counterstained using DAPI (blue). Negative controls were obtained by omission of primary antibodies (pictures in Fig. 46 appendix page 164). Signals were scanned with a confocal laser scanning microscope. Maximum intensity pictures of confocal stacks are shown. All cases show THBS4 positivity in the tumor stroma, either intracellular or extracellular, in between the cytokeratin-positive tumor cells. Zoom-in areas are depicted in white dashed lines. Three representative tumor regions are depicted.

Colocalization with THBS4 could be detected for two of the fibroblast markers, namely vimentin and α-smooth muscle actin. In all analyzed samples, THBS4-positive cells were also positive for these two proteins, as illustrated by yellow coloring in the merge pictures (Fig. 24 & 25). Colocalization with vimentin indicates that THBS4 expressing cells are of mesenchymal nature, whereas colocalization with α-smooth muscle actin further constricts the entity of these cells to the myofibroblast phenotype.

Colocalization of THBS4 and the fibroblast marker procollagen 1 was basically not encountered. In some very rare regions, marginal costaining could be observed (Fig. 25; white arrowheads). Although no clear colocalization could be identified, cells positive for either THBS4 or procollagen 1 were arranged truly close to one another. Tumor regions rich in THBS4-positive cells contained high amounts of procollagen 1-positive cells as well.

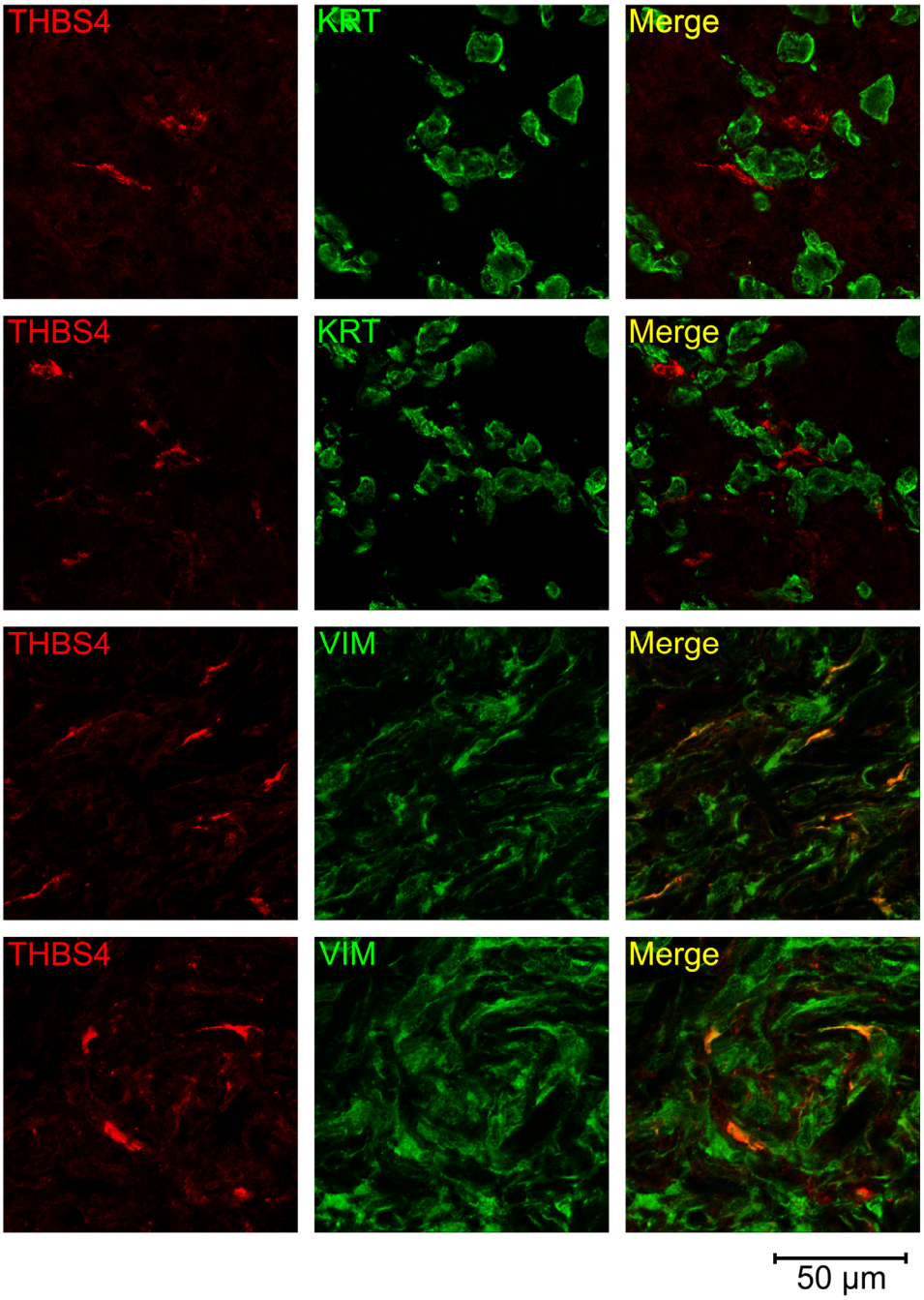


Figure 24: Colocalization of THBS4 and cytokeratin, and THBS4 and vimentin in human diffuse-type gastric adenocarcinomas.

Simultaneous fluorescent immunohistochemical detection of THBS4 in red and either pan cytokeratin (KRT4, 5, 6, 8 10, 13 and 18) or vimentin (VIM) in green was performed on 10 µm thin cryosections. Negative controls were obtained by omission of primary antibodies (pictures in Fig. 46 appendix page 164). Signals were scanned with a confocal laser scanning microscope. Representative pictures of single confocal sections are shown, respectively.

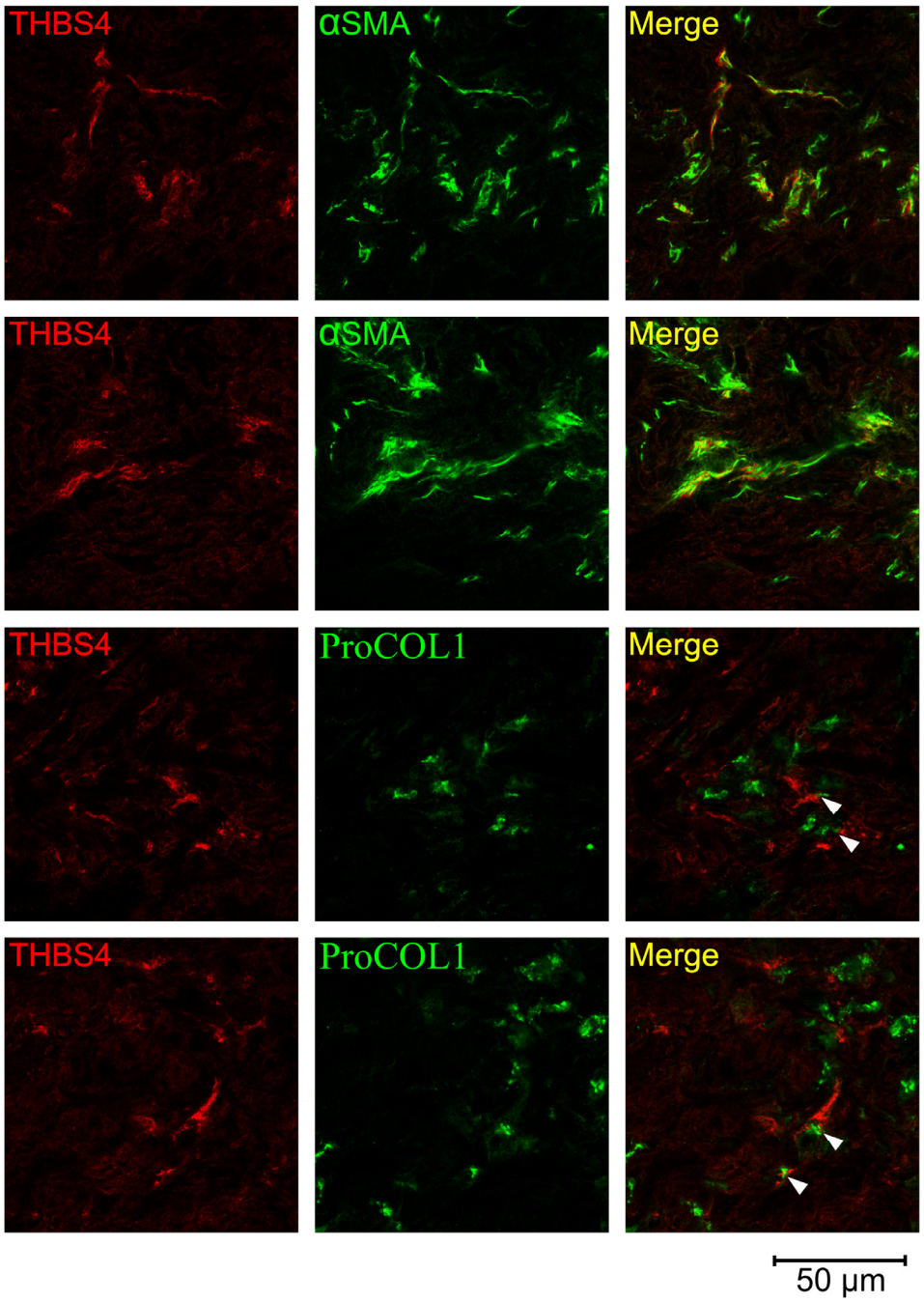


Figure 25: Colocalization of THBS4 and *α*-smooth muscle actin, and THBS4 and procollagen 1 in human diffuse-type gastric adenocarcinomas.

Simultaneous fluorescent immunohistochemical detection of THBS4 in red and either *α*-smooth muscle actin (*α*SMA) or procollagen 1 (ProCOL1) in green was performed on 10 µm thin cryosections. Negative controls were obtained by omission of primary antibodies (pictures in Fig. 46 appendix page 164). Signals were scanned with a confocal laser scanning microscope. Representative pictures of single confocal sections are shown, respectively. White arrowheads mark regions of marginal colocalization.

As a second approach to pinpoint the cellular origin of extracellular THBS4 in human diffuse-type gastric tumors, a panel of different human cell lines, including carcinoma cells of the stomach and of other organ entities, cells from other malignancies and two cell lines from healthy/normal tissue, was examined for *THBS4* mRNA expression. This analysis confirmed that carcinoma cells, regardless of derivation, do not express *THBS4* mRNA. Only one carcinoma cell line, the esophageal carcinoma-derived Kyse-520, possessed some expression, though to a very weak extent. In cells from other malignancies, weak expression levels could be detected in SH-SY5Y, a neuroblastoma cell line; Daudi, a lymphoma cell line, and HL-60, a leukemia cell line. The highest *THBS4* mRNA levels could be identified in a cell line derived from normal embryonic kidney, designated HEK-293. The only fibroblast cell line in this study, the normal forehead skin-derived 142 BR, did not show any relevant *THBS4* mRNA expression (see Fig. 38 appendix page 157).

3.3.2.6 *THBS4* expression in cell lines of diffuse-type gastric cancer-associated fibroblasts and "normal" gastric fibroblasts

The performed immunohistochemical experiments showed that subpopulations of cancer-associated fibroblasts are the cells expressing THBS4 in diffuse-type gastric tumors. However, they do not allow any quantitative comparisons to normal gastric fibroblasts.

Thus, fibroblast cell lines derived from within human diffuse-type gastric adenocarcinomas and from human healthy gastric mucosa were obtained (Department of Surgical Oncology, Osaka City University Medical School, Japan) and analyzed for *THBS4* mRNA abundance via quantitative real-time PCR. Matched pairs of these cancer-associated (CAFs) and normal fibroblasts (NFs) were available from two patients.

This analysis revealed that CAFs from each pair possess significantly (p<0.01, Mann-Whitney-U-test) higher *THBS4* mRNA amounts than NF counterparts (Fig. 26). The fold change of overexpression in CAFs compared to NFs yielded ≈2 in patient pair 32 and ≈3.3 in pair 33. However, this effect was most prominent within the first passages of cell culture and diminished over the time course of culture. These decreasing differences in *THBS4* expression were mainly driven by shrinking expression levels CAFs. The expression intensity in NFs remained fairly unchanged (data not shown). At passages 12 to 14 the differential expression of *THBS4* had basically disappeared and CAFs and NFs exhibited similar amounts of *THBS4* mRNA.

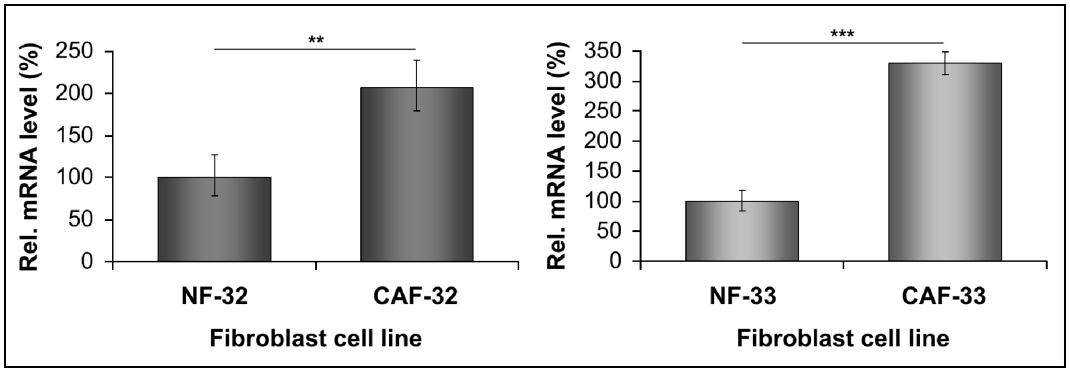


Figure 26: *THBS4* mRNA expression in human diffuse gastric cancer-associated fibroblasts and normal gastric fibroblasts.

The mRNA abundance of *THBS4* was examined in matched pairs of cancer-associated and normal fibroblasts by means of quantitative real-time PCR in triplicates. All cell lines were at passages 4–6 when analyzed. Quantitation was done relative to the transcript of *GAPDH* and expression levels in normal fibroblasts were set to 100%, respectively. Significance in differential expression between groups was calculated using Mann-Whitney-U-test and p-value of two-tailed asymptotic significance was chosen as significance estimate. Error bars represent integrated standard errors of the mean.

NF – normal fibroblast; CAF – cancer-associated fibroblast; Rel. – relative; ** – p<0.01; *** – p<0.001

3.3.2.7 *In vitro* analysis of tumor cell-dependent activation of gastric fibroblasts and accompanying differences in *THBS4* expression

In order to demonstrate that increased *THBS4* expression in cancer-associated fibroblast of diffuse-type gastric tumors is a result of fibroblast activation by tumor cells, an *in vitro* activation approach using indirect coculture via conditioned medium was chosen.

Different human gastric fibroblast cell lines were treated with the conditioned medium of different human diffuse-type gastric adenocarcinoma-derived tumor cell lines. The fresh tumor cell line-specific medium was used as control medium. After incubation with either tumor cell-conditioned or control medium, fibroblasts were harvested and lysed, followed by extraction of total RNA. Quantitative real-time PCRs were performed to measure *THBS4* mRNA expression levels, respectively.

Two diffuse type-derived tumor cell lines (OCUM-2M and OCUM-8) were employed in these experiments to elucidate whether tumors own potential differences in the ability and potency to activate fibroblasts and stimulate *THBS4* expression. Both diffuse gastric cancer-associated fibroblasts (CAFs) and normal gastric fibroblasts (NFs) were challenged with the tumor cell-conditioned medium, respectively, to uncover potential differences in capability to become activated. Fibroblast pair NF/CAF-33 was used for these studies.

Stimulation with the conditioned medium of both tumor cell lines investigated significantly ($p<0.01$, Mann-Whitney-U-test) increased *THBS4* expression in NF-33 cells. This increase was about 2.5 fold when using the conditioned medium of OCUM-2M and about 1.5 fold for OCUM-8. In CAF-33, increases in *THBS4* mRNA abundance upon stimulation with the conditioned media were encountered, too. However, these increases were slightly less significant ($p<0.1$, Mann-Whitney-U-test) as compared to the ones observed in the NF counterpart. The

elevation in *THBS4* mRNA levels in CAF-33 was about 1.25 fold for OCUM-2M and 1.75 for OCUM-8, respectively (Fig. 27).

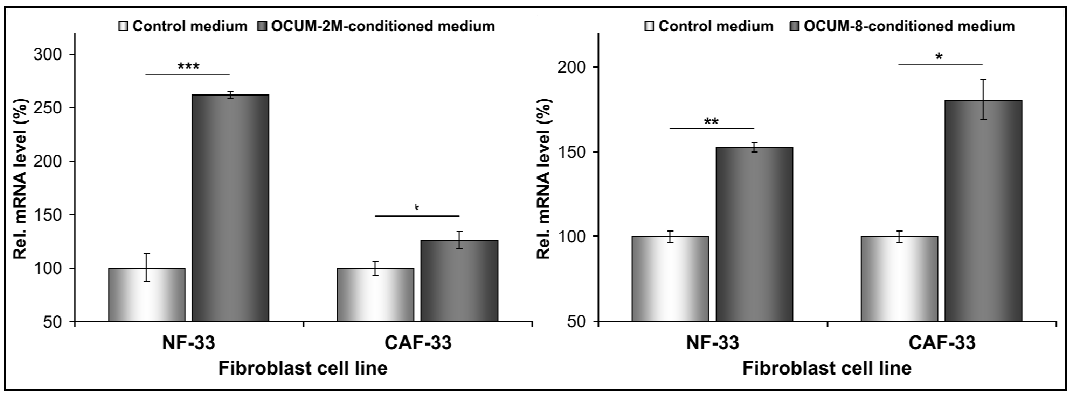


Figure 27: *THBS4* mRNA expression in human diffuse gastric cancer-associated fibroblasts and normal gastric fibroblasts upon stimulation with tumor cell-conditioned medium.

Matched pairs of cancer-associated and normal fibroblasts were incubated for 48 hours with conditioned medium derived from either human diffuse-type gastric tumor cell line OCUM-2M or OCUM-8. Fresh OCUM-2M and OCUM-8-specific medium was used as control medium, respectively. All experiments were run in triplicates. The *THBS4* mRNA abundance was examined by means of quantitative real-time PCR in triplicates. Quantitation was done relative to the transcript of *GAPDH* and expression levels in control medium-treated fibroblasts were set to 100%, respectively. Significance in differential expression between groups was calculated using Mann-Whitney-U-test and *p*-value of two-tailed asymptotic significance was chosen as significance estimate. Error bars represent integrated standard errors of the mean. Results of representative experiments are shown.

NF – normal fibroblast; CAF – cancer-associated fibroblast; Rel. – relative; ½* – $p<0.1$; * – $p<0.05$; ** – $p<0.01$; *** – $p<0.001$

This microarray-based gene expression profiling study identified THBS4 as the most potent marker for histological type of gastric adenocarcinoma. Strong transcriptional and protein overexpression is present in diffuse-type tumors compared to intestinal ones. In diffuse-type tumors, THBS4 is expressed and secreted by cancer-associated fibroblasts upon stimulation by tumor cells and accumulates in the extracellular matrix of the tumor stroma. In conclusion, THBS4 represents a highly abundant extracellular constituent of the activated tumor stroma of diffuse-type gastric adenocarcinomas.

3.4 Identification of prognostic gene signatures and marker genes for N+ gastric adenocarcinomas

3.4.1 Extraction of prognostic candidate genes from microarray data

Identification of candidate genes with putative prognostic value was performed separately for diffuse and intestinal-type gastric adenocarcinoma patients. One of the 59 patients originally enrolled in the microarray study was excluded due to inappropriate patient data (refer to chapter 2.1.1). Thus, 32 intestinal and 26 diffuse patients were used for the identification of prognostic candidates.

Diffuse and intestinal-type patients were grouped according their disease outcome. Patients that did not show formation of postoperative metastases or locoregional recurrence of disease within at least 3.6 years after R0 resection were categorized as patients with "good prognosis", whereas patients with recurrent disease were grouped as possessing "poor prognosis". Statistical significance testing comparing each gene's average expression in the good prognosis group to its average in the poor prognosis group (Welch-test) was performed to extract genes featuring differential expression between the prognostic groups. Correction for false-positives via multiple testing procedures (Bonferroni FWER, Benjamini and Hochberg FDR *etc*.) resulted in no significant features passing the test for both histological patient cohorts. Hence, no *p*-value correction could be implemented in these statistical tests. Because of missing multiple testing correction, the *p*-value cutoff at which probe sets/genes were considered to be of significant relevance was chosen to be as low as possible, namely <0.001 (non-corrected). The fold change cutoff was chosen to be 1.5 (Fig. 28).

For intestinal-type patients, 36 probe sets covering 35 unique transcripts fulfilled the given criteria (Fig. 28; Tab. 19 appendix page 175 ff.). Covered transcripts comprised 24 annotated sequences, with all of them being fully characterized as protein coding, and 11 non-annotated sequences (orphan open reading frames, hypothetical loci, *etc*.; NetAffx, September 2009). The majority of genes were identified to be upregulated within the patient subgroup with poor prognosis, namely 31, whereas only 4 were upregulated in the patients with good prognosis. Gene Ontology analysis for genes upregulated/downregulated in either prognostic group did not reveal any significantly enriched GO terms (data not shown). Manual screening of single genes elucidated that 6 of the 31 genes upregulated in the patient subgroup with poor prognosis are somehow involved in transcriptional processes, such as DNA/RNA/nucleic acid binding or transcription factor activity (*HOXC10*, *MLLT3*, *ZNF266*, *MLF1*, *DDIT3/NR1H3* and *SOX7*). Thus, a potential biological connection with regard to transcription might be present within these genes, despite "negative" results of GO analysis.

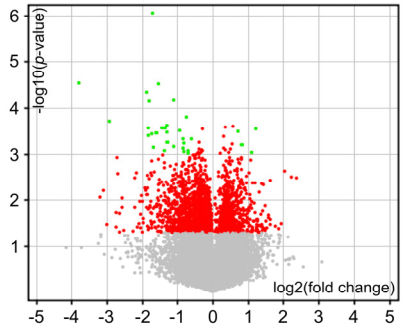
For diffuse-type patients, 18 probe sets covering 14 unique annotated transcripts and 1 unique non-annotated one could be identified as being of putative prognostic value (Fig. 28; Tab. 20 appendix page 179 ff.). Among annotated sequences, 13 had been characterized as being protein coding (NetAffx, September 2009). Identically to the intestinal cohort, most of the identified genes were upregulated in the patient group with poor prognosis, namely 10, whereas as little as 5 genes showed overexpression in the patients with good prognosis. Gene Ontology analysis for genes upregulated/downregulated in either prognostic group did also not reveal any significantly enriched GO terms (data not shown).

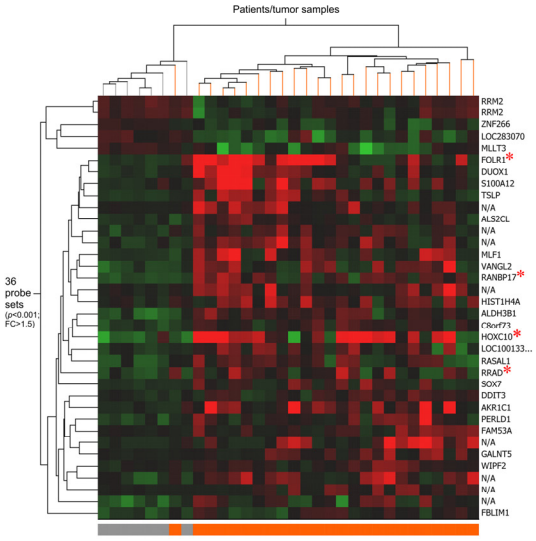
No overlapping members of the two gene signatures could be encountered.

Intestinal-type patients

Recurrent *vs.* non-recurrent disease

Welch-test	*p* all	*p* >0.05	*p* <0.05	*p* <0.02	*p* <0.01	*p* <0.005	*p* <0.001
FC all	54,675	52,301	2,374	956	495	250	52
FC>1.1	20,499	18,395	2,104	892	480	249	52
FC>1.5	2,561	1,786	775	395	230	135	36
FC>2.0	629	334	295	162	99	66	24
FC>3.0	138	61	77	50	34	24	10
expected by chance			2,733	1,093	546	273	54

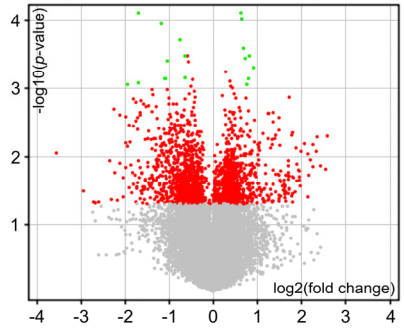


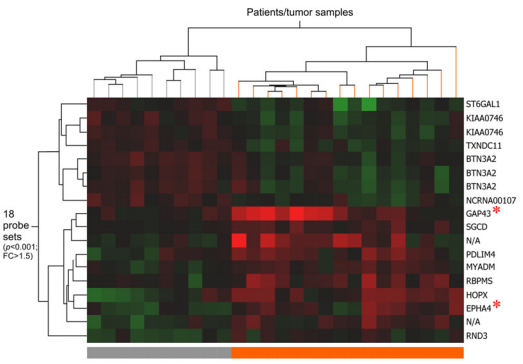


Diffuse-type patients

Recurrent *vs.* non-recurrent disease

Welch-test	*p* all	*p* >0.05	*p* <0.05	*p* <0.02	*p* <0.01	*p* <0.005	*p* <0.001
FC all	54,675	52,747	1,928	726	340	158	26
FC>1.1	20,541	18,722	1,819	713	338	158	26
FC>1.5	2,678	1,889	789	372	209	103	18
FC>2.0	710	404	306	164	96	46	7
FC>3.0	181	77	104	57	31	19	3
expected by chance			2,733	1,093	546	273	54





for figure legend see next page

Figure 28: Schematic workflow of identification of prognostic candidate genes for N+ gastric adenocarcinoma patients from microarray data.

Identification of candidate genes possessing putative prognostic value was performed separately for diffuse and intestinal-type patients. Patients were grouped according to recurrence of disease, respectively, and Welch-test was employed to extract significantly differentially expressed genes. Numbers of probe sets passing different *p*-value cutoff filters and fold change cutoffs are summarized in tables. Volcano plots visualize all probe sets according to *p*-value and fold change, respectively. Color-coding of probe sets is equal in depicted tables and volcano plots (green: $p<0.001$, FC>1.5; red: $p<0.05$, all FCs; grey: $p>0.05$; all FCs). Two-way hierarchical clustering heatmap visualizes the normalized expression data of significantly differentially expressed genes ($p<0.001$, FC>1.5), respectively. Genes marked with asterisks were chosen for validation via quantitative real-time PCR.

N+ – regional lymph node-positive; *vs.* – *versus*; FC – fold change

3.4.2 Validation of prognostic candidate genes via quantitative real-time PCR

From the two panels of identified candidate genes comprising putative prognostic value, only genes overexpressed in the patients with poor prognosis were regarded as suitable candidates for future applications in clinical routine and thus chosen for validation. For the validation procedure, additional seven patients were added to the original cohort of 58 patients, on whose basis the prognostic candidate genes had been identified. These additional patients had initially been excluded from the patient cohort of the microarray study, because respective tumor samples contained less than 50% tumor cells.

For intestinal-type patients, ras-related associated with diabetes (*RRAD*), homeobox C10 (*HOXC10*), RAN binding protein 17 (*RANBP17*) and folate receptor 1 (*FOLR1*) were selected for validation. *RANBP17* was chosen, because it exhibited the highest significance in differential expression (p=8.77E-07, Welch-test), whereas *HOXC10* was selected, because it showed the strongest upregulation (fold change ~14) within the poor prognosis group. *RRAD* and *FOLR1* were chosen due to their literature record: RRAD had been identified to regulate growth and tumorigenicity of breast cancer cells *in vitro* and *in vivo* and to interact with the putative tumor metastasis suppressor NME1 (non-metastatic cells 1, protein

expressed in), also known as NM23 (Tseng *et al*. 2001, Zhu *et al*. 1999). FOLR1 overexpression had been associated with poor prognosis of uterine adenocarcinomas (Allard *et al*. 2007). However, in non-small-cell lung cancers, higher levels of *FOLR1* appeared to be correlated with more favorable prognosis (Iwakiri *et al*. 2008).

The growth associated protein 43 (*GAP43*) and ephrin receptor A4 (*EPHA4*) were chosen for validation for the diffuse-type patient cohort. *EPHA4* was selected, because a prognostic potential had been proposed in previous studies: In gastric cancer, EPHA4 overexpression had been associated with recurrence and shortened survival (Oki *et al*. 2008); and in colorectal cancers, *EPHA4* overexpression had been correlated with liver metastases (Oshima *et al*. 2008). *GAP43* was selected, because it showed the strongest upregulation (fold change ~3.8) within the poor prognosis group.

Relative mRNA expression levels of candidates were determined by means of quantitative real-time PCR. Usefulness and accuracy of measured relative mRNA amounts to distinguish the recurrent patients from the non-recurrent ones was examined using the receiver operating characteristic, subsequently. These analyses were performed on the initial patient cohort of 58 patients and additionally on the expanded cohort of 65 patients. The area under the curve, the 95% confidence interval and the *p*-value resulting from these analyses are summarized in Table 9.

According to this PCR analysis, *RANBP17*, *FOLR1*, *RRAD* and *HOXC10* own the capability to significantly separate the recurrent from the non-recurrent intestinal-type patients of this study. The strongest power could be identified for *RANBP17*, indeed. In contrast, *GAP43* and *EPHA4* did not show any explicit power in discriminating the respective prognostic groups of diffuse-type patients. However, *GAP43* possessed a tendency for significance. Interestingly, *RRAD*, one of the candidates originally selected for the intestinal cohort, showed a tendency to separate recurrent from non-recurrent diffuse-type patients of this study as well.

In general, significances obtained on the expanded cohort were slightly stronger compared to the ones obtained on the initial cohort. Thus, inclusion of mentioned additional patients with tumor samples of tumor contents below 50% does not weaken the prognostic performance of identified candidate genes. Hence, all following analyses were performed on the enlarged cohort of 65 patients.

Table 9: Evaluation of candidate gene's mRNA expression to discriminate recurrent from non-recurrent N+ gastric adenocarcinoma patients.

Receiver operating characteristics, including calculation of area under the curve (of sensitivity against 1-specificity), 95% confidence interval and *p*-value, were performed on relative mRNA expression levels as determined by quantitative real-time PCR. The analyses were conducted separately for diffuse and intestinal-type patients. Orange-colored *p*-values are <0.1 and *p*-values depicted in red are <0.05.

recurr. – recurrence; *vs.* – *versus*; diff. – diffuse; int. – intestinal; AUC – area under the curve; CI – confidence interval

58 patients (used for microarray analysis)

	Recurr. *vs.* no recurr. in **diff.** type Receiver operating characteristic			Recurr. *vs.* no recurr. in **int.** type Receiver operating characteristic		
Transcript	AUC	95% CI	*p*-value	AUC	95% CI	*p*-value
GAP43	0.706	0.504–0.900	0.082	0.535	0.314–0.760	0.767
EPHA4	0.575	0.339–0.811	0.527	0.474	0.203–0.745	0.837
RANBP17	0.525	0.293–0.757	0.833	0.840	0.702–0.978	0.007
FOLR1	0.344	0.126–0.561	0.188	0.766	0.601–0.930	0.034
RRAD	0.686	0.491–0.881	0.092	0.754	0.585–0.924	0.043
HOXC10	0.650	0.406–0.894	0.206	0.757	0.591–0.923	0.040

65 patients (microarray plus additional)

	Recurr. *vs.* no recurr. in **diff.** type Receiver operating characteristic			Recurr. *vs.* no recurr. in **int.** type Receiver operating characteristic		
Transcript	AUC	95% CI	*p*-value	AUC	95% CI	*p*-value
GAP43	0.706	0.516–0.896	0.063	0.591	0.385–0.798	0.460
EPHA4	0.574	0.355–0.792	0.507	0.478	0.201–0.754	0.857
RANBP17	0.510	0.293–0.727	0.929	0.857	0.734–0.980	0.004
FOLR1	0.333	0.131–0.535	0.132	0.773	0.615–0.932	0.027
RRAD	0.686	0.491–0.881	0.092	0.754	0.594–0.913	0.040
HOXC10	0.583	0.360–0.807	0.452	0.776	0.620–0.932	0.025

Subsequently, a more detailed analysis of the prognostic value was performed for each candidate (Fig. 29 & 30). This included, as the first step, the examination of the distribution of relative mRNA expression levels in the recurrent and non-

recurrent patient groups via box and whisker plot. Secondly, an appropriate cutoff point of relative mRNA levels distinguishing between recurrent and non-recurrent patients was selected from the receiver operating characteristics curve. According to this cutoff point, patients were subsequently divided into patients possessing "high" or "low" mRNA levels. A finally performed survival analysis using the Kaplan-Meier method was used to elucidate whether these expression-based patient subgroups possess differences in the length of survival until relapse or progression of disease (disease-free survival).

In concordance to the results of receiver operating characteristics, *RANBP17*, *FOLR1*, *RRAD* and *HOXC10* were able to significantly distinguish intestinal-type patients with shorter disease-free survival from the ones with more favorable prognosis in Kaplan-Meier analysis. Again, *RANBP17* exhibited the strongest significance in doing so. In contrast, *GAP43* and *EPHA4* failed to significantly separate these different prognostic groups in diffuse-type patients, respectively. However, *RRAD* showed again a significant separation of the diffuse-type prognostic groups. Thus, the prognostic value of this gene was also examined for the joined cohort of both histological types (Fig. 31).

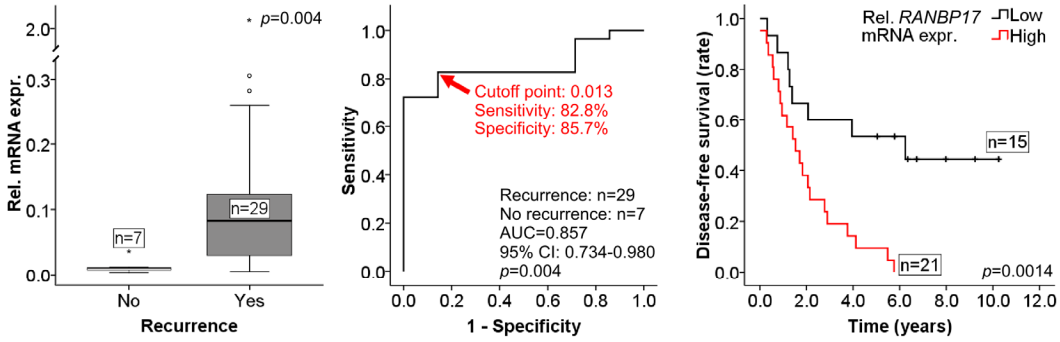
for figure see next page

Figure 29: Evaluation of prognostic value of *RANBP17* (A), *FOLR1* (B), *RRAD* (C) and *HOXC10* (D) mRNA expression for N+ intestinal-type gastric adenocarcinoma patients.

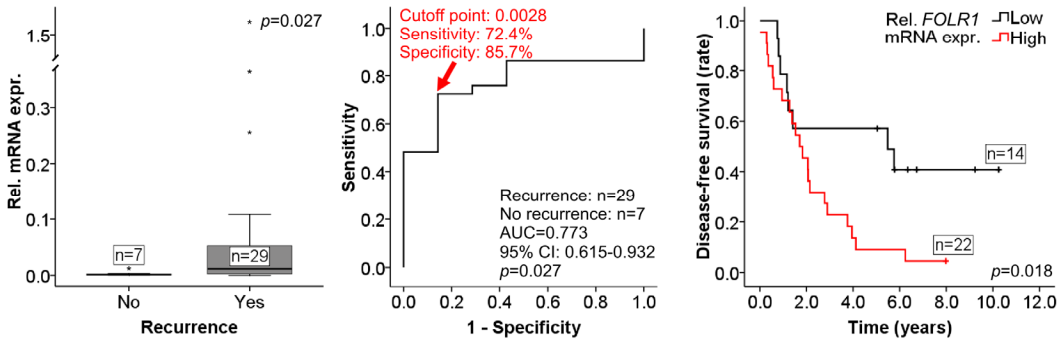
Levels of mRNA abundance were measured by means of quantitative real-time PCR. Quantitation was done relative to the transcript of *β*-actin. Distribution of relative mRNA amounts in recurrent and non-recurrent patients was examined using box and whisker plots (left panel). Significance in differential mRNA expression was determined by Mann-Whitney-U-test and *p*-value of two-tailed asymptotic significance was chosen as significance estimate. Receiver operating characteristics were performed to select appropriate cutoff points of mRNA expression, respectively (middle panel). Selected cutoffs are marked by red arrows and sensitivity and specificity of the test system at that point are indicated. For survival analyses patients were dichotomized according to selected cutoffs and survival curves were plotted according to the Kaplan-Meier method (ticks indicate censored data). Significance in survival difference between groups was assessed by logrank test.

rel. – relative; expr. – expression; n – number; AUC – area under the curve; CI – confidence interval

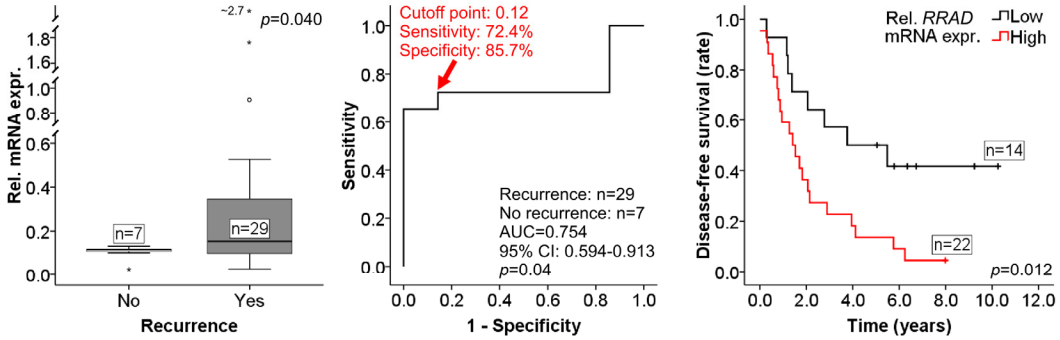
A: *RANBP17*



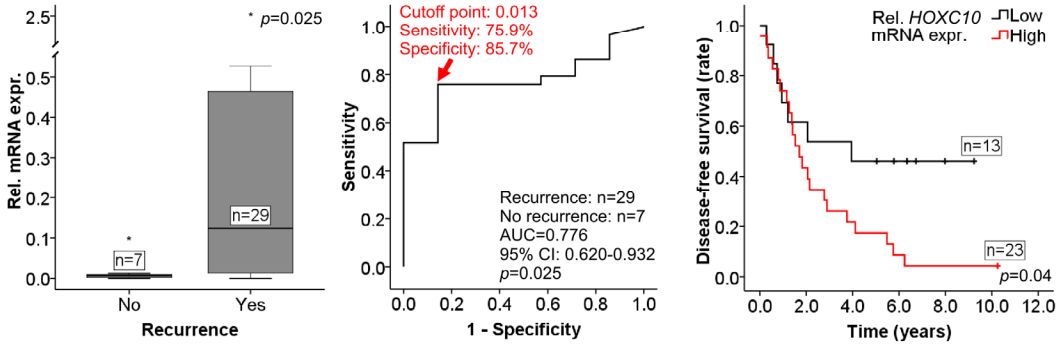
B: *FOLR1*



C: *RRAD*



D: *HOXC10*



for figure legend see previous page

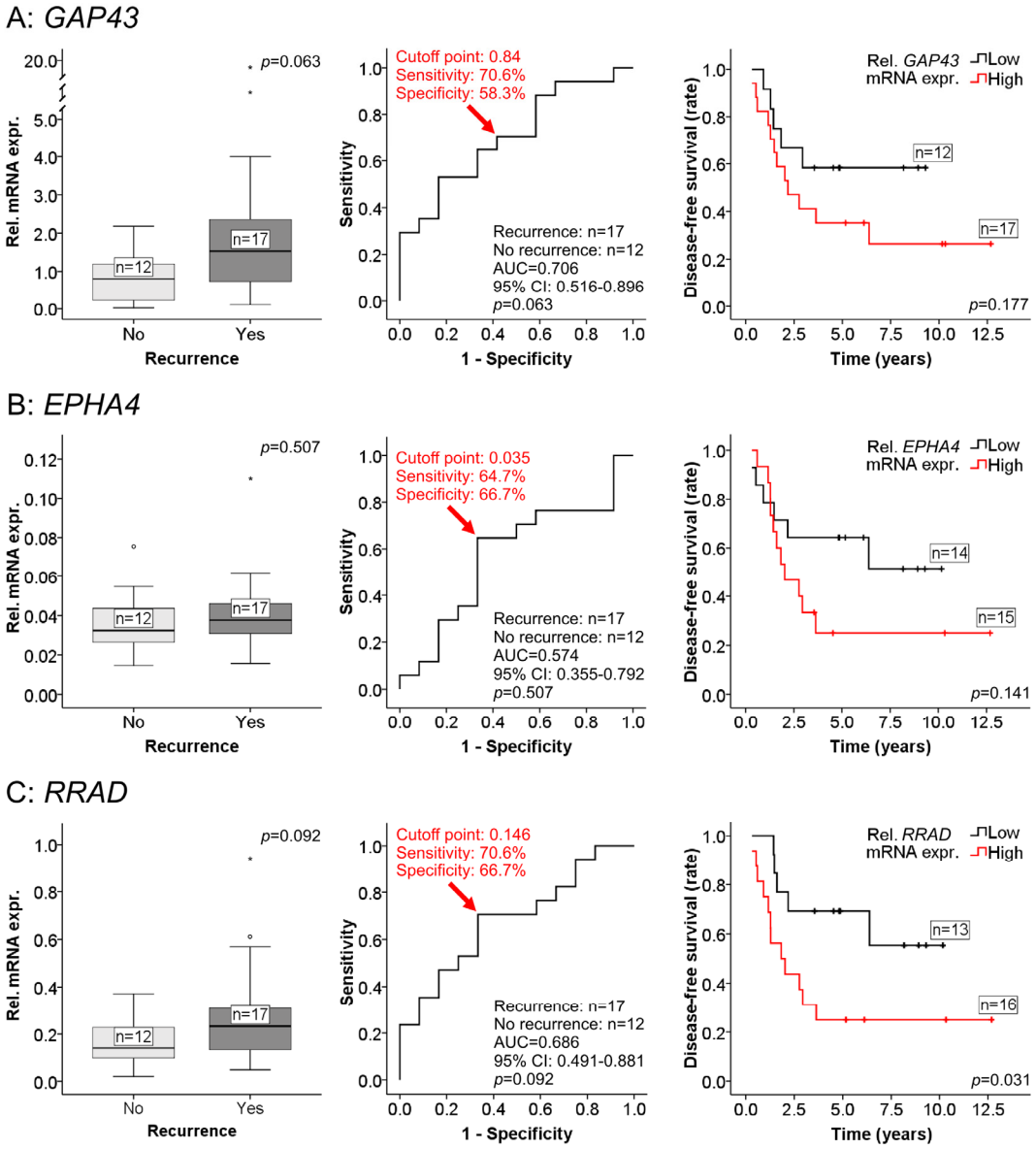


Figure 30: Evaluation of prognostic value of *GAP43* (A), *EPHA4* (B) and *RRAD* (C) mRNA expression for N+ diffuse-type gastric adenocarcinoma patients.

Levels of mRNA abundance were measured by means of quantitative real-time PCR. Quantitation was done relative to the transcript of *β*-actin. Distribution of relative mRNA amounts in recurrent and non-recurrent patients was examined using box and whisker plots (left panel). Significance in differential mRNA expression was determined by Mann-Whitney-U-test and *p*-value of two-tailed asymptotic significance was chosen as significance estimate. Receiver operating characteristics were performed to select appropriate cutoff points of mRNA expression, respectively (middle panel). Selected cutoffs are marked by red arrows and sensitivity and specificity of the test system at that point are indicated. For survival analyses, patients were dichotomized according to selected cutoffs and survival curves were plotted according to the Kaplan-Meier method (ticks indicate censored data). Significance in survival difference between groups was assessed by logrank test.

rel. – relative; expr. – expression; n – number; AUC – area under the curve; CI – confidence interval

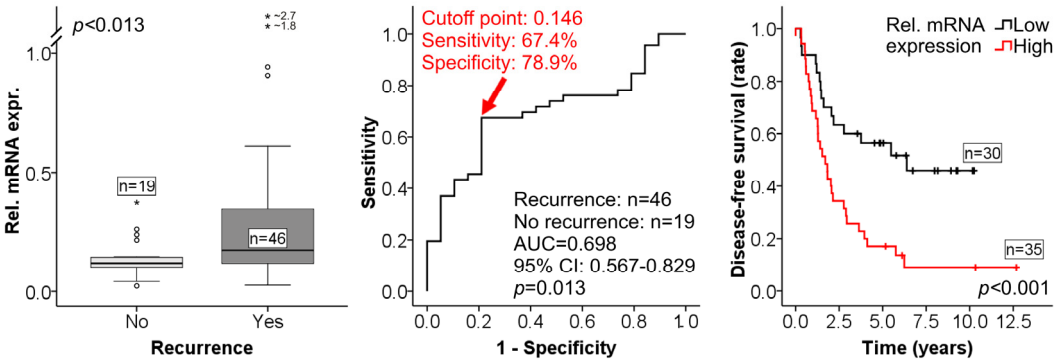


Figure 31: Evaluation of prognostic value of *RRAD* mRNA expression for human N+ gastric adenocarcinoma patients (of all histological types).

Levels of mRNA abundance were measured by means of quantitative real-time PCR. Quantitation was done relative to the transcript of *β*-actin. Distribution of relative mRNA amounts in recurrent and non-recurrent patients was examined using box and whisker plots (left panel). Significance in differential mRNA expression was determined by Mann-Whitney-U-test and *p*-value of two-tailed asymptotic significance was chosen as significance estimate. Receiver operating characteristics were performed to select appropriate cutoff points of mRNA expression, respectively (middle panel). Selected cutoffs are marked by red arrows and sensitivity and specificity of the test system at that point are indicated. For survival analyses, patients were dichotomized according to selected cutoffs and survival curves were plotted according to the Kaplan-Meier method (ticks indicate censored data). Significance in survival difference between groups was assessed by logrank test.

rel. – relative; expr. – expression; n – number; AUC – area under the curve; CI – confidence interval

Only if a marker carries “unique” information that is non-redundant to the information offered by other factors, it bears potential benefits for clinical routine. Thus, the independence of the prognostic candidates from other conventional prognostic factors, such as T-stage or N-stage, needs to be evaluated. Multivariate analyses, which are the methods of choice regarding this problem, could only be conducted in the intestinal cohort. The diffuse one contained insufficient patient numbers. This analysis clearly showed that all candidate genes selected for the intestinal patients offer prognostic information that is independent of T-stage and N-stage (Tab. 10). Thus, they own “novel” non-redundant information regarding prognosis.

Table 10: Results of multivariate analyses with respect to recurrence of disease of N+ gastric adenocarcinoma patients.

Multivariate analyses were performed based on binary logistic regression. The analysis was performed separately for each gene with N-stage and T-stage being introduced as covariates, respectively. The gene's relative mRNA expression levels were introduced dichotomized as "high" and "low" with the same cutoff points used as in Figure 29. Depicted are *p*-values from the tests with significant ones being colored in red.

vs. – *versus*

Parameter	**Univariate**	**Multivariate**			
		(*RANBP17*)	**(*FOLR1*)**	**(*RRAD*)**	**(*HOXC10*)**
N-stage (1 *vs.* 2)	0.051	0.998	0.998	0.998	0.998
T-stage (1/2 *vs.* 3/4)	0.434	0.329	0.212	0.500	0.641
RANBP17 ("high" *vs.* "low")	<0.001	<0.005	/	/	/
FOLR1 ("high" *vs.* "low")	<0.005	/	0.022	/	/
RRAD ("high" *vs.* "low")	<0.005	/	/	0.009	/
HOXC10 ("high" *vs.* "low")	0.224	/	/	/	0.019

Additionally, the correlation of candidate gene's mRNA expression with the most important prognostic clinicopathological parameters, namely N-stage, T-stage and UICC-stage, was examined via non-parametric statistical tests. This analysis unraveled that the candidate genes basically do not correlate with these parameters. Slightly significant correlation with N-stage and UICC-stage could only be observed for one gene, namely *HOXC10* (Tab. 11 & 12; *p*-values <0.05 are depicted in red). Tendencies for significance were present for *RRAD* for T-stage and for *GAP43* and *EPHA4* for N-stage (Tab. 11 & 12; *p*-values <0.1 are depicted in orange). However, all these weak correlations were encountered in the intestinal patient cohort, only.

Table 11: Correlation of prognostic gene's mRNA expression with clinicopathological parameters in N+ intestinal-type gastric adenocarcinoma patients.

Relative mRNA levels of prognostic genes were evaluated for differential expression between subgroups of clinicopathological parameters by means of Mann-Whitney-U-test or Kruskal-Wallis-H-test. The table summarizes the *p*-values obtained from the tests. Values depicted in orange show a tendency for significance ($p<0.1$). For these comparisons, the direction of regulation is indicated with asterisks marking upregulation in advanced stages and open circles specifying downregulation in advanced stages.
Data on detailed composition of clinicopathological groups and subgroups can be obtained from Tab. 2, page 33. For UICC-stage, a simplified version with stages IIIA and IIIB being combined to one central stage III was used.

T – depth of penetration of primary tumor; N – metastatic involvement of regional lymph nodes; UICC – *Union Internationale Contre Le Cancer*; *vs.* – *versus*

Parameter	*GAP43*	*EPHA4*	*RANBP17*	*FOLR1*	*RRAD*	*HOXC10*
T-stage (1–4)	0.179	0.986	0.560	0.404	0.137	0.186
T-stage (1/2 *vs.* 3/4)	0.183	0.810	0.268	0.936	0.098*	0.047*
N-stage (1 *vs.* 2)	0.089*	0.057*	0.986	0.619	0.279	0.959
UICC-stage (I–IV)	0.415	0.757	0.552	0.417	0.407	0.197
UICC-stage (I/II *vs.* III/IV)	0.856	0.365	0.256	0.934	0.420	0.038*

Table 12: Correlation of prognostic gene's mRNA expression with clinicopathological parameters in N+ diffuse-type gastric adenocarcinoma patients.

Relative mRNA levels of prognostic genes were evaluated for differential expression between subgroups of clinicopathological parameters by means of Mann-Whitney-U-test or Kruskal-Wallis-H-test. The table summarizes the *p*-values obtained from the tests. Values depicted in red possess significance ($p<0.05$). Values depicted in orange show a tendency for significance ($p<0.1$). For these comparisons, the direction of regulation is indicated with asterisks marking upregulation in advanced stages and open circles specifying downregulation in advanced stages.
Data on detailed composition of clinicopathological groups and subgroups can be obtained from Tab. 2, page 33). N3-patients were excluded from analysis due to lack of sufficient number. For UICC-stage, a simplified version with stages IIIA and IIIB being combined to one central stage III was used.

T – depth of penetration of primary tumor; N – metastatic involvement of regional lymph nodes; UICC – *Union Internationale Contre Le Cancer*; *vs.* – *versus*

Parameter	*GAP43*	*EPHA4*	*RANBP17*	*FOLR1*	*RRAD*	*HOXC10*
T-stage (1–4)	0.352	0.392	0.684	0.594	0.104	0.856
T-stage (1/2 *vs.* 3/4)	0.281	0.529	0.753	0.621	0.022	0.556
N-stage (1 *vs.* 2)	0.581	0.209	0.292	0.880	0.422	0.288
UICC-stage (I–IV)	0.459	0.322	0.750	0.635	0.220	0.631
UICC-stage (I/II *vs.* III/IV)	0.893	0.590	0.686	0.822	0.106	0.222

Besides analysis in gastric adenocarcinoma tissue, mRNA expression of identified prognostic candidate genes was also examined in a panel of different human cell lines (see Fig. 39–44 appendix page 158 ff.). This panel included a broad selection

of human carcinoma cells (gastrointestinal tract and others), cells from other human malignancies (lymphomas, leukemia, melanoma, neuroblastoma) and two cell lines from normal human tissue (embryonic kidney epithelial cells and forehead skin fibroblasts). This analysis uncovered that none of the investigated genes showed relevant mRNA amounts in all carcinoma cell lines of this panel. Noteworthy mRNA expression could only be observed in subpopulations or even just single carcinoma cell lines, with *RRAD* and *GAP43* carrying the most limited expression profile. In contrast, *HOXC10* and *RANBP17* exhibited the most consistent expression level across analyzed carcinoma cells, with *HOXC10* showing an especially universal expression in gastric carcinoma cells of the panel.

This microarray-based gene expression profiling study identified two putative gene signatures for prognosis of N+ gastric adenocarcinomas – one for intestinal-type patients and one for diffuse-type ones. However, the significance in correlation with prognosis was stronger for the members of the signature for intestinal-type patients. Individual candidate genes were selected for validation via quantitative real-time PCR. Transcriptional *RANBP17*, *FOLR1* and *HOXC10* expression were validated to significantly stratify the intestinal-type patients of this study with respect to disease-free survival. *RANBP17* showed the strongest capability in doing so. Transcriptional *GAP43* and *EPHA4* expression showed tendencies for disease-free survival-based stratification of diffuse-type patients of this study. One transcript, designated *RRAD*, was validated to be prognosis-relevant for both histological cohorts of this study, indeed.

3.4.3 Evaluation of BMP and activin membrane-bound inhibitor homolog as a prognostic gene for N+ gastric adenocarcinomas

The BMP and activin membrane-bound inhibitor homolog (BAMBI) had recently been identified to be linked to a potentially aggressive tumor phenotype and predict tumor recurrence and metastatic potential of colorectal cancer (Fritzmann *et al*. 2009, Togo *et al*. 2008).

In the data set of this microarray study, the probe set covering the *BAMBI* transcript was found to exhibit significantly higher expression values within diffuse-type patients with recurrence of disease (patients with "poor prognosis") compared to the non-recurrent (patients with "good prognosis") ones (p=0.00356, Welch-test; FC=3.74). However, no statistical significance (p=0.74, Welch-test; FC=1.27) could be detected for the comparison of these groups in intestinal-type patients of this study.

Because of its postulated prognostic value for colorectal cancer and the identified significant association with prognosis of a patient subgroup of this cohort, *BAMBI* was chosen as an additional candidate gene for validation via quantitative real-time PCR.

Foremost, the quantitative real-time PCR analysis revealed that *BAMBI* mRNA is in general much more abundant in intestinal than in diffuse-type gastric adenocarcinomas. Only few diffuse samples possess comparable amounts as present in intestinal type (Fig. 32).

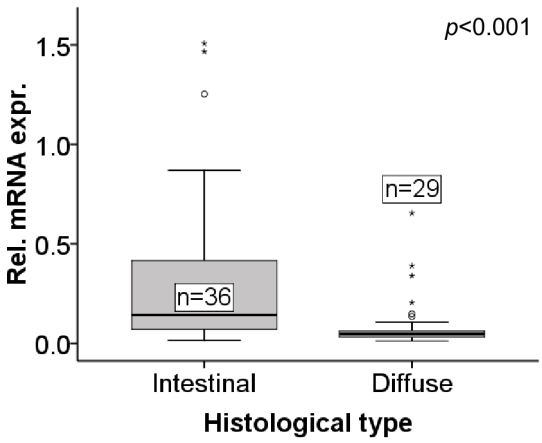


Figure 32: *BAMBI* mRNA expression in human N+ diffuse and intestinal-type gastric adenocarcinomas.

The mRNA abundance of *BAMBI* was examined by means of quantitative real-time PCR. Quantification was done relative to the transcript of *β*-actin. Significance in differential expression between groups was calculated using Mann-Whitney-U-test and *p*-value of two-tailed asymptotic significance was chosen as significance estimate. Expression value distribution within groups is displayed by box and whisker plots.

Rel. – relative; expr. – expression; n – number

Since one of the latest publications concerning BAMBI had described it to be a marker for the metastatic potential of colorectal cancers (Fritzmann *et al*. 2009), *BAMBI* mRNA expression was not just examined for correlation with recurrence in general, but also, more specifically, for correlation with absence or presence of postoperative metastasis formation. However, the evaluation concerning both outcome types individually was only possibly for the intestinal cohort of this study. Three of the intestinal patients with recurrence of disease had faced locoregional recurrence without formation of metastases. Thus, these patients could be regrouped according to postoperative metastasis status. In the diffuse-type population of this study, recurrence of disease was always accompanied by formation of metastases, meaning that locoregional recurrence was never encountered alone, without formation of any metastases. The correlation analysis was performed using the receiver operating characteristic including calculation of area under the curve (of sensitivity against 1-specificity), 95% confidence interval and *p*-value. It revealed that in the diffuse cohort, *BAMBI* mRNA expression does slightly correlate with recurrence of disease/formation of postoperative metastases. However, only a marginal trend towards statistical significance was present for this correlation (p=0.121, Mann-Whitney-U-Test). For the study population of intestinal-type patients, no such correlation with outcome related parameters could be observed. Nevertheless, the observed significances were faintly stronger when considering

postoperative metastasis formation alone (Tab. 13). Thus, all subsequently performed analyses were conducted with respect to formation of postoperative metastases.

Table 13: Evaluation of *BAMBI* mRNA expression to discriminate recurrent/postoperatively metastasizing from non-recurrent/non-postoperatively metastasizing N+ gastric adenocarcinoma patients.

Receiver operating characteristics, including calculation of area under the curve (of sensitivity against 1-specificity), 95% confidence interval and *p*-value, were performed on relative mRNA expression levels as determined by quantitative real-time PCR. The analyses were conducted separately for diffuse and intestinal-type patients and also for the 58 patient cohort with tumor samples of >50% tumor content and the 65 patient cohort including additional samples of <50% tumorous tissue. Orange-colored *p*-values are <0.1.

ROC – receiver operating characteristic; postop. – postoperative; AUC – area under the curve; CI – confidence interval; n – number

		Diffuse patients	Intestinal patients	
		Recurrence/postop. metastases	Recurrence	Postop. metastases
ROC (n=58)	AUC	0.712	0.571	0.573
	95% CI	0.514–0.911	0.325–0.818	0.344–0.802
	p-value	0.073	0.569	0.515
ROC (n=65)	AUC	0.672	0.567	0.569
	95% CI	0.467–0.876	0.325–0.808	0.347–0.792
	p-value	0.121	0.589	0.525

Despite non-significant receiver operating characteristic regarding dichotomized outcome status, subsequently performed survival analyses clearly showed that patients possessing “high” or “low” *BAMBI* mRNA expression levels exhibit significant differences in metastasis-free survival. This difference could be observed for both histological patient cohorts, with a stronger significance present for diffuse-type patients (Fig. 33).

Additionally, the correlation of *BAMBI* mRNA expression with the most important prognostic clinicopathological parameters, namely N-stage, T-stage and UICC-stage, was examined via non-parametric statistical tests. Except for N-stage

of the diffuse cohort, no correlation with any of the examined clinicopathological parameters could be observed in this analysis (Tab. 14).

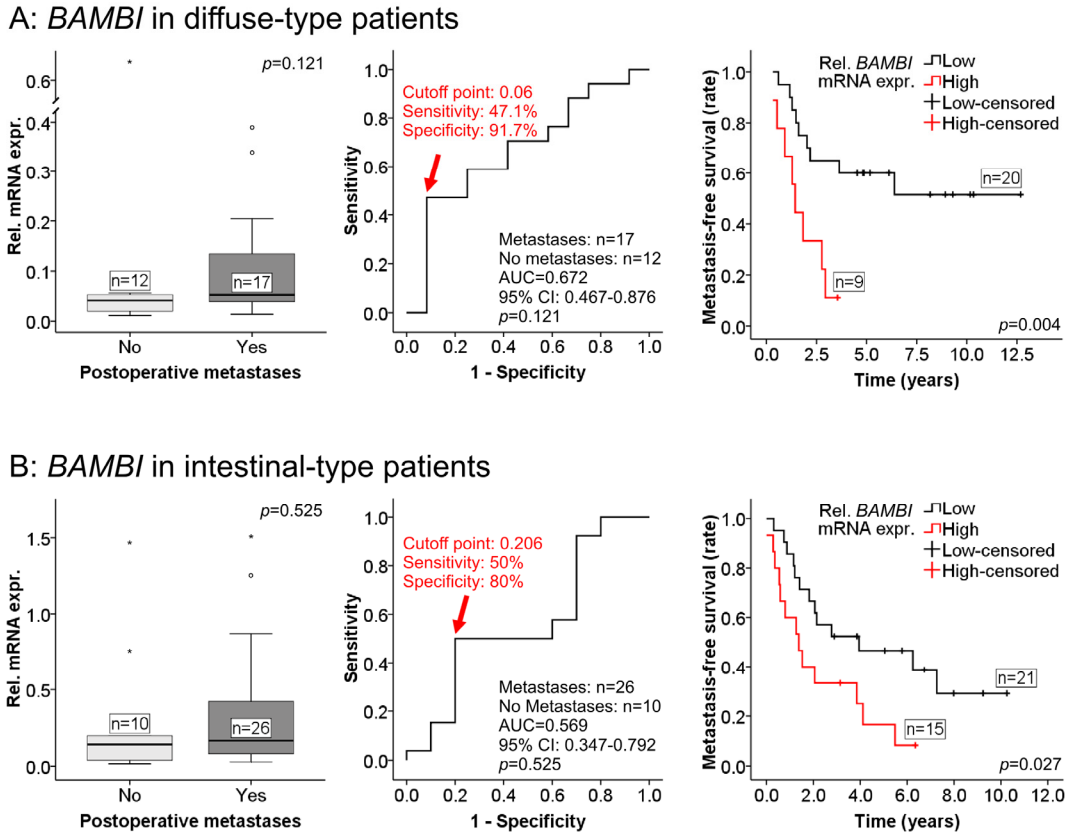


Figure 33: Evaluation of prognostic value of *BAMBI* mRNA expression for N+ gastric adenocarcinoma patients.

Analysis was performed separately for diffuse and intestinal-type patients. Levels of mRNA abundance were measured by means of quantitative real-time PCR. Quantitation was done relative to the transcript of *β*-actin. Distribution of relative mRNA amounts in postoperatively metastasized and non-metastasized patients was examined using box and whisker plots (left panel). Significance in differential mRNA expression was determined by Mann-Whitney-U-test and *p*-value of two-tailed asymptotic significance was chosen as significance estimate. Receiver operating characteristics were performed to select appropriate cutoff points of mRNA expression, respectively (middle panel). Selected cutoffs are marked by red arrows and sensitivity and specificity of the test system at that point are indicated. For survival analyses, patients were dichotomized according to selected cutoffs and survival curves were plotted according to the Kaplan-Meier method. Significance in survival difference between groups was assessed by logrank test.

Rel. – relative; expr. – expression; n – number; AUC – area under the curve; CI – confidence interval; metastases – postoperative metastases

Table 14: Correlation of *BAMBI* mRNA expression with selected clinicopathological parameters in human N+ diffuse and intestinal-type gastric adenocarcinomas.

Relative mRNA levels were evaluated for differential expression between subgroups of clinicopathological parameters by means of Mann-Whitney-U-test or Kruskal-Wallis-H-test. Summarized are the *p*-values obtained from the tests. Values depicted in red possess significance ($p<0.05$). Data on detailed composition of clinicopathological groups and subgroups can be obtained from Tab. 2, page 33. N3-patients were excluded from analysis due to lack of sufficient number. For UICC-stage, a simplified version with stages IIIA and IIIB combined to one central stage III was used. Asterisks indicate upregulation in higher stages.

pat. – patients; T – depth of penetration of primary tumor; N – metastatic involvement of regional lymph nodes; UICC – *Union Internationale Contre Le Cancer*

Parameter	**Diffuse pat.**	**Intestinal pat.**
	p-value	*p*-value
T-stage (1–4)	0.577	0.590
T-stage (1/2 *vs*. 3/4)	0.418	0.205
N-stage (1 *vs*. 2)	0.027*	0.770
UICC-stage (I–IV)	0.428	0.681
UICC-stage (I/II *vs*. III/IV)	0.192	0.300

In order to localize *BAMBI* mRNA expression within diffuse and intestinal-type gastric adenocarcinomas, *in situ* hybridization experiments were performed. These experiments clearly assigned *BAMBI* mRNA to the tumor cells in both histological types and ascertained the already identified differences in abundance (Fig. 34). In intestinal-type tumors, *BAMBI* mRNA could be detected in basically all cells of the discrete tumor with rather equal intensities, whereas in diffuse samples only single tumor cells, usually growing in Indian file pattern, showed positivity for the transcript. These single tumor cells tended to be present at regions of invasion into healthy aspects, *e.g.* the muscularis, whereas the tumor mass itself appeared to be negative for *BAMBI* mRNA (Fig. 34 E). The observed intensity of stained tumor cells within both histological types did not seem to differ (Fig. 34 A/B *vs.* C/D). No positivity for the *BAMBI* transcript could be observed in epithelial cells (Fig. 34 F/G) or any other cell entity (smooth muscle cells, endothelial cells, stromal cells *etc*.; data not shown) of non-neoplastic gastric mucosa.

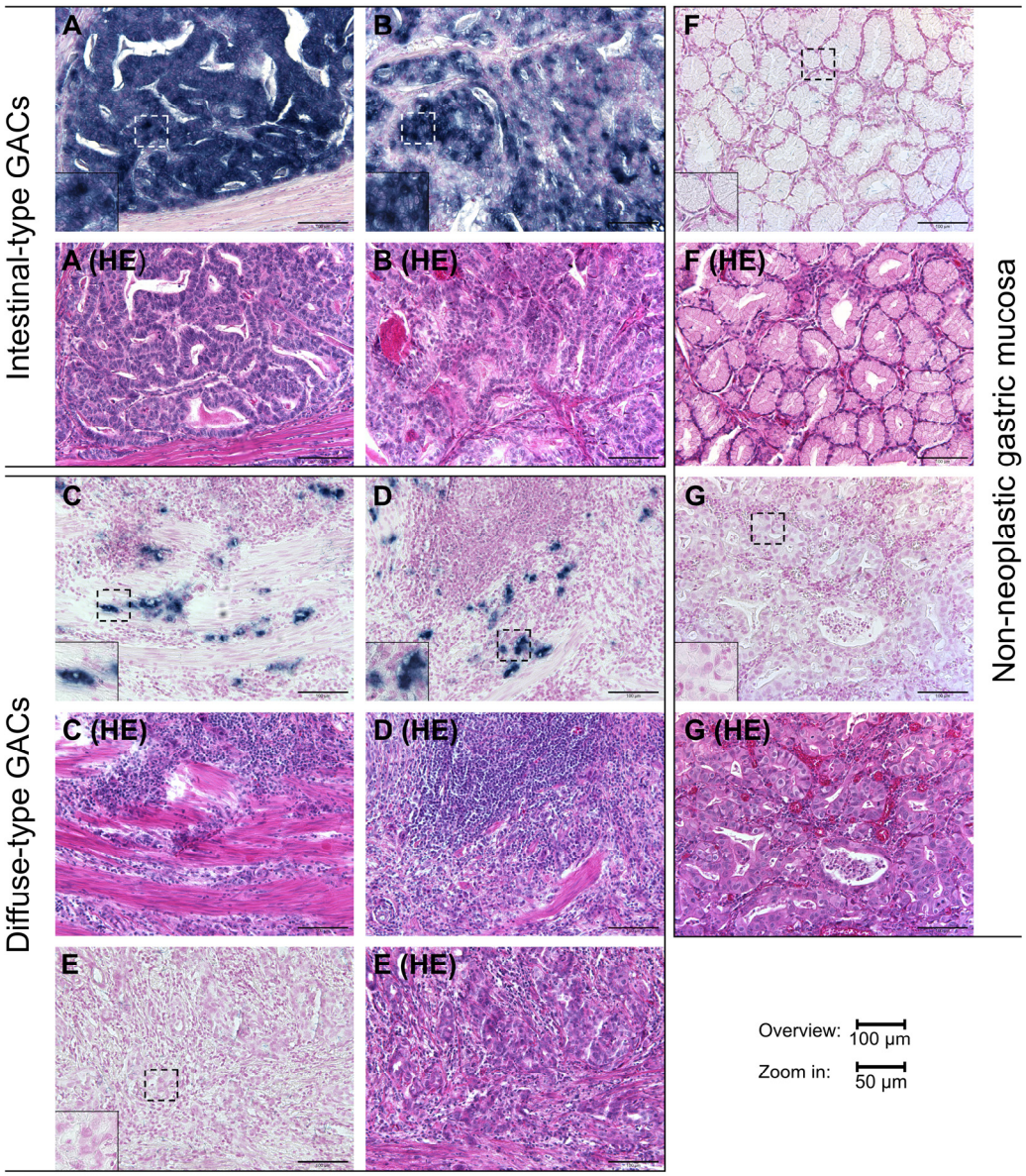


Figure 34: *BAMBI* mRNA expression in human gastric adenocarcinomas.

In situ hybridizations of *BAMBI* mRNA were performed on 10 µm thin formalin-fixed paraffin-embedded tissue sections. *BAMBI* mRNA is stained in dark purple/blue using BM Purple. Cell nuclei are counterstained using Nuclear Fast Red. Representative sections of intestinal and diffuse-type tumors and non-neoplastic mucosa are depicted in 200× magnification with zoom-in areas (dashed lines) in 400× magnification. Tumor cells in HE-pictures can be recognized by irregular size and rather big nuclei with irregular coarse chromatin and hyperchromatism. Note how the tumor cells of the whole tumor mass in intestinal-type samples are positive for *BAMBI* mRNA (A & B), whereas only few tumor cells of diffuse samples exhibit positivity (C & D). *BAMBI*-positive diffuse-type tumor cells were predominantly encountered within regions of invasion (C & D; single tumor cells invading the muscularis), whereas the tumor itself usually did not possess any *BAMBI*-positivity (E). Non-neoplastic gastric mucosa did not contain any relevant levels of *BAMBI* mRNA (F & G).

HE – hematoxylin-eosin (-staining)

Besides analysis in gastric adenocarcinoma tissue, *BAMBI* mRNA expression was also examined in a panel of different human cell lines (see Fig. 45 appendix page 164). This panel included a broad selection of human carcinoma cells (gastrointestinal tract and others), cells from other human malignancies (lymphomas, leukemia, melanoma, neuroblastoma) and two cell lines from normal human tissue (embryonic kidney epithelial cells and forehead skin fibroblasts). According to this analysis, *BAMBI* is expressed by the vast majority of malignant cell lines, though to varying extends. Expression was also encountered in the two normal cell lines under investigation. However, the expression in carcinoma cells was in general stronger than the one identified in the normal cell lines.

This study showed that the prognostic value of *BAMBI* mRNA expression, which had been recently identified for colorectal cancers, can be expanded to N+ gastric adenocarcinoma patients. *BAMBI* mRNA expression was of significant correlation with metastasis-free survival time of patients, regardless of histological type of adenocarcinoma. However, the mRNA expression profile was identified to differ between the two histological types. Intestinal-type tumors showed a rather consistent expression across the whole tumor aspect, whereas in diffuse-type samples only single tumor cells predominant at regions of invasion showed expression.

4 Discussion

4.1 Global gene expression profiles of diffuse and intestinal-type gastric adenocarcinomas

Gene expression profiling of the whole human genome was performed on transcriptional level using the Affymetrix microarray technique. Non-microdissected tissue samples (whole specimens) were used for this study to allow more global description of the entire aspect of the tumor scenario, instead of focusing on the tumor cells, only. This approach was driven by recent advances in cancer research that had revealed a strong relevance of tumor-stroma interactions via extracellular matrix molecules, matrix metalloproteinases, angiogenic factors *etc*. for cancer progression (Arias 2001, Matrisian *et al*. 2001).

The entire gene expression profiling procedure was comprehensively controlled. Only untreated cryomaterial was used for RNA extraction; and only RNAs of high quality/integrity (RIN>7; refer to chapter 2.2.3) were further processed. The amplification and labeling process was controlled according to the guidelines of the "GeneChip® Expression Analysis" technical manual (Affymetrix Inc., Santa Clara, USA). Quality of microarray data was controlled before data analysis, once in-house and additionally by the company of MicroDiscovery (MicroDiscovery GmbH, Berlin, Germany).

In this study, the molecular distinctiveness of diffuse and intestinal adenocarcinomas became already evident by unsupervised hierarchical clustering of samples, in which the two types generally clustered apart from one another with few exceptions. This observation indicates that the vast majority of the genome must be differentially expressed between them and strengthens the idea that they represent

distinct tumor entities with strikingly different molecular backgrounds (Chan *et al*. 1999).

The general separation of the two histological types on the basis of unsupervised approaches had previously been encountered by other expression profiling studies on gastric cancer as well (Boussioutas *et al*. 2003, Jinawath *et al*. 2004). However, expression profiling studies failing to distinguish the two types on the basis of unsupervised procedures exist, too (Hippo *et al*. 2002). Some expression profiling studies obtained two major gastric tumor subclusters in unsupervised clustering, but data on histological type is missing in these studies (Chen *et al*. 2003, Leung *et al*. 2004). However, it is very likely that the two clusters in these studies represent the histological groups, respectively. The observation that "misgrouped" samples of either histological type exist had also been encountered before (Boussioutas *et al*. 2003).

Consequentially, statistical significance testing comparing each gene's average expression in the diffuse-type group to its average in the intestinal-type group identified many hundreds of genes as differentially expressed between the types. Depending on the multiple testing procedure used for correction for false-positives, different numbers of genes passing the test were encountered, with at least ~250 being truly differentially expressed as determined by the Bonferroni FWER. Many other gene expression profiling studies on gastric cancer had been able to identify differentially expressed genes between these two histological groups, with some of them even generating classifier genes or signatures (Boussioutas *et al*. 2003, Hippo *et al*. 2002, Jinawath *et al*. 2004, Norsett *et al*. 2004, Wu *et al*. 2006). Thus, in concordance with other studies, this work demonstrates that a gene expression profile-based stratification of the two histological types of gastric adenocarcinoma as described by Laurén is possible and useful, and that the phenotypic differences are reflections of broad molecular distinctiveness.

However, a huge difference exists between the number of differentially expressed genes identified in this study and those of previous gene expression profiling studies on gastric cancer. Numbers of differentially expressed genes observed in previous studies range from 15 over 27 to 46 (Hippo *et al*. 2002, Jinawath *et al*. 2004, Wu *et al*. 2006), whereas this study identified at least ~250 (according to Bonferroni FWER) and at maximum ~1,500 (according to Benjamini and Hochberg FDR) genes. Thus, of all data sets currently available, the number of genes identified as differentially expressed is by far highest in the here presented study. The most obvious reason for this is that this study was conducted using a microarray that covers more sequences, namely the whole human transcriptome, than all the ones used in previous investigations. All previous studies had been performed using arrays covering subsets of the transcriptome, only, *e.g.* customized arrays. Other reasons may include differences in sample selection, extents of contamination with non-tumor cells, sample preparation, RNA quality and performance of the microarray platform, in general.

In principle, this study comprises all the selection steps necessary for a successful microarray study offering reliable data. These steps include employment of untreated cryo-preserved tissue, control for minimal amounts of non-tumor cells within the analyzed samples, RNA quality check prior to microarray processing, general quality control of microarray data and application of multiple testing procedures for statistical tests. All previous microarray-based gene expression profiling studies on gastric cancer have certain drawbacks regarding one or multiple of these parameters. Thus, this study represents the first one to gene expression-profile gastric cancer with respect to histological type on the most global and comprehensively controlled level possible to date.

In order to gain insight into the biology underlying each histological type and its behavior, Gene Ontology (GO) analysis was performed. This analysis was conducted using a statistical approach for identification of significant GO-term en-

richment implemented in the freely available software package GOSSIP (Bluthgen *et al*. 2005). It unraveled that genes overexpressed in either type are connected to different biological processes with no overlap in significantly enriched GO terms. For genes overexpressed in intestinal-type tumors, a predominant association with proliferation and growth-connected processes, such as cell cycle and mitosis could be encountered. In contrast, most of the genes overexpressed in diffuse-type tumors encode for proteins of the extracellular matrix and/or for proteins that play important roles in adhesion or developmental processes. This is fully consistent with others studies, which likewise identified genes of the extracellular matrix, cell adhesion, cell migration and cell-matrix-interaction to be upregulated in diffuse-type gastric tumors (Jinawath *et al*. 2004) and genes related to cell proliferation and enhancement of cell growth (DNA replication, spindle assembly, chromosome segregation) to be upregulated in the intestinal type (Boussioutas *et al*. 2003, Jinawath *et al*. 2004). Nevertheless, this work is the first to underline these strong differences in biology by using a GO-based statistical approach.

The results of GO-analysis strengthen once more that diffuse and intestinal-type gastric adenocarcinomas represent distinct diseases with unique biological and molecular features (Chan *et al*. 1999).

4.2 Thrombospondin 4 – the most potent marker for histological type of gastric adenocarcinoma in this data set

In the data set of this microarray-based gene expression profiling study, the thrombospondin 4 (*THBS4*) transcript was the one possessing the strongest correlation with histological type of gastric adenocarcinoma. It was vastly overexpressed in the diffuse type compared to the intestinal one and represented the most potent marker for discerning both types.

THBS4 is a secreted multidomain glycoprotein of the extracellular matrix belonging to a family of at least 5 thrombospondins. Each THBS4 molecule is composed of the conserved heparin-binding THBS N-terminus, four EGF-like domains, with two of them being calcium binding, eight THBS type-III repeats, a predicted cell attachment site and the conserved THBS C-terminus (Fig. 35). As a mature form, five THBS4 molecules assemble to form a homopentamer, which can be detected by electron microscopy. The molecular weight of monomeric THBS4 is ~106 kDa, whereas the pentamer holds ~550 kDa (Lawler *et al*. 1995).

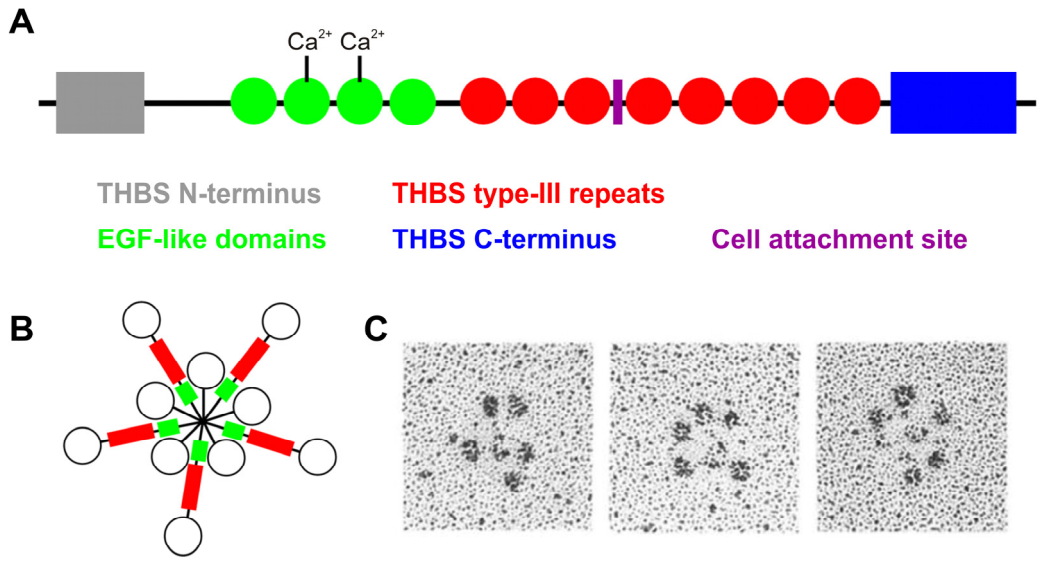


Figure 35: Human THBS4 protein structure.

A: The protein and its termini, conserved domains and repeats and the potential cell attachment site are depicted. Two of the EGF-like domains are predicted to be calcium-binding. All information was obtained from UniProtKB/Swiss-Prot (state of December 2009). B: As a mature form, five THBS4 molecules assemble to form a pentamer with globular regions at the ends (open circles) and EGF-like domains and type-III repeats in the middle (schematic view adopted from (Lawler *et al*. 1995). C: Electron microscopy pictures of purified recombinant THBS4 (Lawler *et al*. 1995).

THBS – thrombospondin; EGF – epidermal growth factor

Thrombospondins play an important role in a variety of processes that involve tissue remodeling including embryonic development, wound healing, angiogene-

sis, synaptogenesis, neoplasia and tumor progression (Kazerounian *et al*. 2008). However, most of the information about this protein family relies on data generated for THBS1 and THBS2, the two best-characterized family members. Comparably little is known about the other three family members.

Quantitative real-time PCRs clearly validated the strong overexpression of *THBS4* in the diffuse-type tumors of this study and signified that expression is principally absent from the intestinal type. Interestingly, the amounts of *THBS4* expression detected in the two diffuse-type samples, which had been "misgrouped" to the intestinal cluster in microarray clustering analysis, ranked right in between the amounts observed in either histological group. Expression was not completely undetectable like in the intestinal samples and was slightly lower as the ones observed in the diffuse population. Thus, PCR analysis proved that these diffuse-type samples can in fact still be assigned to their correct histological group based on *THBS4* expression. Here, one of the major problems of the microarray technique becomes obvious, namely the lack of producing true quantitative data and the limited sensitivity of measurements, which leads to a poor ability to detect subtle differences (Evans *et al*. 2003), such as the ones encountered here.

Future studies need to elucidate whether the here postulated marker property of *THBS4* for histological type can be globally expanded to gastric adenocarcinomas or whether it represents a cohort-specific artifact. Such future investigations include the validation of the observed differential expression in other patient cohorts of gastric cancer, so-called independent test sets, as a key step.

Analysis and validation of genes of interest in independent data sets has been facilitated by online microarray data banks that have recently emerged. A data base including three major microarray-based gene expression profiling data sets on gastric cancer was recently established by the Korean Research Institute of Biosciences & Biotechnology (Kim *et al*. 2007b). According to this data base, *THBS4* mRNA is differentially expressed between the two histological types in at least two other studies, namely the ones by Chen *et al.* and Boussioutas *et al.*. Never-

theless, one of the studies comprised in this data base failed to identify a differential expression, namely the one by Hippo *et al.* (Tab. 15). However, the Hippo *et al.* study has one major drawback, which is that only five diffuse gastric tumor samples were introduced in the analysis, whereas 17 intestinal samples were employed. In principle, the data obtained from this microarray data collection indicates that the differential expression of *THBS4* observed in this study is not solely a cohort-specific event, but can be expanded to other gastric cancer populations in addition.

Table 15: Differential *THBS4* mRNA expression with respect to histological type of human gastric adenocarcinoma in three independent microarray data sets in comparison to this study.

Data was obtained from http://human-genome.kribb.re.kr/stomach/ (Kim *et al*. 2007b), an online service allowing the evaluation of single genes in the data of three major microarray studies on gastric cancer (Boussioutas *et al*. 2003, Chen *et al*. 2003, Hippo *et al*. 2002) and their meta-analysis. Statistical significance in differential expression between diffuse and intestinal-type tumors was estimated by *t*-test. Calculation of *q*-values from the corresponding *p*-values was performed to control for the FDR that occurs in multiple hypothesis testing (Storey and Tibshirani 2003). Meta-analysis of the data sets was performed as follows: the meta fold change values were calculated by taking an arithmetic mean of the individual values and meta *p* and *q*-values were calculated by taking the arithmetic mean of log(*p*-values) and log(*q*-values) of the individual data sets. The *q*-value of *THBS4* for this study was calculated using QVALUE software (Storey 2002, Storey and Tibshirani 2003) from the non-corrected *p*-value (*p*-values of all genes imputed). *p*-values depicted in red possess statistical significance ($p<0.05$).

FDR − False Discovery Rate; FC − fold change; diff. − diffuse; int. − intestinal

Diff.:int.	Chen	Boussioutas	Hippo	Meta	This study
FC (log2)	1.4383	0.9556	1.1227	1.1722	5.35
p-value	1.95E-05	0.0008	0.1949	0.0001	4.66E-12
q-value	0.0031	0.0195	0.8083	0.0078	8.19E-08

In published data of other gene expression profiling studies that had aimed at identifying genes differentially expressed between diffuse and intestinal-type gastric adenocarcinomas or at establishing robust classifiers, THBS4 does not appear in any of the published gene lists (Jinawath *et al*. 2004, Norsett *et al*. 2004, Wu *et al*. 2006). However, this does not necessarily mean that it is not differentially ex-

pressed, but that it was either not among the genes with the highest significance or could simply not be detected, because the microarray used did not contain probes for its transcript sequence. The latter may be because; these studies used so-called customized in-house arrays, which contain subselections of the transcriptome, only. Unfortunately, no data indicating whether the *THBS4* transcript was covered or not is available for these studies.

Regarding future validation of *THBS4* in independent patient cohorts, special attention needs to be drawn to lymph node-negative gastric adenocarcinoma patients. Since the gene had been identified on the basis of lymph node-positive patients, only, it is of major interest to elucidate whether the *THBS4* overexpression in an event exclusive for diffuse-type gastric tumors of node-positive patients or applies to node-negative patients, too. However, some evidence points to the fact that this phenomenon can in fact be expanded to node-negative patients. For example, the study of Boussioutas *et al.*, which found *THBS4* to possess differential expression in the same sense, included also node-negative patients. The study of Chen *et al.* does not contain any information on N-stage or any other clinicopathological parameter routinely assessed.

Immunohistochemical studies clearly confirmed that *THBS4* mRNA overexpression in diffuse-type gastric adenocarcinomas results in protein overexpression as well. No THBS4 protein could be detected in the intestinal-type population examined. In healthy gastric epithelium and connective tissue of mucosa and submucosa, the sites from which carcinoma formation originates and first progresses into, no THBS4 expression was visible, too. Thus, THBS4 expression in diffuse-type tumors is not a remainder of the healthy counterparts, but a tumor-specific feature. Given these information, the encountered transcriptional marker property can be expanded to protein level as well.

The main localization of THBS4 expression within diffuse-type gastric adenocarcinomas was demonstrated to be the extracellular matrix of the tumor stroma. The tumor stroma is the so-called microenvironment of a tumor and comprises extracellular matrix and embedded cells (macrophages, endothelial cells, fibroblasts, *etc.*). In malignant tumors the stroma can constitute up to 90% of the tumor mass. Its characteristics are strongly modulated by the tumor in an entity-specific manner and in return exploited by the tumor for progression, metastasization *etc.* (Liotta and Kohn 2001). Since especially strong THBS4 expression was observed in regions of high tumor cells density and invasion, the protein might be linked to proliferative or invasive properties of the tumor. In immunohistochemical colocalization studies, the cells expressing and secreting THBS4 were identified to be cancer-associated fibroblasts (CAFs) of the myofibroblast phenotype. CAFs are the activated fibroblasts of malignant tumor scenarios and represent one major component of the tumor stroma. Additional terms used to denote them comprise carcinoma-associated fibroblasts, tumor-associated fibroblasts, peritumoral fibroblasts or reactive stromal fibroblasts. Such CAFs have acquired a modified phenotype, similar to fibroblasts found in wound healing. However, to this date it is not fully clear whether CAFs represent a unique fibroblast phenotype or whether they represent a common phenotype of activation upon injury (Kalluri and Zeisberg 2006). Nevertheless, in the last years it has become increasingly evident that CAFs are prominent modifiers and promoters of tumor growth and progression (Kalluri and Zeisberg 2006, Mueller and Fusenig 2004).

The finding that THBS4 is expressed by subpopulations of fibroblasts and not the tumor cells could be further strengthened by screening a cell line panel of different entities for *THBS4* mRNA expression. In this screen, *THBS4* expression was basically not detectable in a wide range of human secondary tumor cell lines including cells from gastric and other carcinoma entities as well as from other malignancies. Since no explicit expression could be detected in the fibroblast line 142BR, *THBS4* expression seems not to be a general attribute of all fibroblasts. Only certain, potentially activated fibroblasts, seem to possess the ability to ex-

press it. HEK-293 cells exhibited the highest *THBS4* mRNA expression levels of all cells investigated. This finding is somewhat difficult to interpret, because the exact cellular origin of HEK cells is still unknown. Embryonic kidney cultures, which are the source of these cells, are composed of a mixture of cell entities including fibroblasts, endothelial cells and epithelial cells. Thus, the true "nature" of HEK cells can be of either one of these entities. Recently, a lot of attributes of developing neurons and neuronal progenitor cells could be assigned to HEK cells. This relationship to neuronal lineage cells is most probably a result of the adenovirus transformation done while generating the cell line (Shaw *et al*. 2002).

To allow quantitative comparison of THBS4 expression in CAFs to normal fibroblasts (NF), *THBS4* mRNA expression was examined in matched pairs of human diffuse gastric cancer-associated fibroblasts and normal gastric fibroblasts. In this analysis, the two examined CAF lines exhibited 2–3.3 fold higher expression levels than the NF counterparts. Since CAFs represent an activated type of fibroblasts, the observed elevated *THBS4* expression levels are most likely a result of fibroblast activation.

In the context of cancer, the fibroblast activation process can most probably be attributed to the tumor cells themselves. In order to investigate the changes in *THBS4* expression in fibroblasts upon stimulation by tumor cells, an indirect coculture approach using the tumor cell-conditioned medium was established. The conditioned medium of tumor cells is depleted of some compounds, which have been partially used up by the tumor cells, but is enriched by tumor cell-specific material, such as secreted proteins (*e.g.* growth factors, enzymes). The conditioned medium of different diffuse-type gastric adenocarcinoma derived tumor cell lines was used to challenge NFs as well as CAFs, thus allowing to model fibroblast activation by tumor cells *in vitro*. Both NFs and CAFs showed elevated levels of *THBS4* expression upon challenge with every tumor cell-conditioned medium. However, depending on the fibroblast cell line as well as the tumor cell line used, different extents of increase were encountered. These experiments di-

rectly prove that elevated *THBS4* expression in fibroblasts is a result of activation by tumor cells. However, additional activation arousing by other cell entities (*e.g.* leukocytes), through reactive oxygen species or complement factor C1, or altered extracellular matrix composition can not be excluded to cause enhanced *THBS4* expression as well (Kalluri and Zeisberg 2006). The fact that not just NFs but also CAFs, which already possess higher initial *THBS4* expression levels, were able to increase expression upon treatment with tumor cell-conditioned media indicates that the initial *THBS4* expression levels in CAFs had not reached the plateau level yet. This would mean in return that the extent of activation regulates the extent of expression in a dose and time-depended manner and not in an on-and-off-switch mode. Repeated exposure to tumor cell-conditioned media and examination of different time-points could give rise to the kinetics underlying *THBS4* expression in fibroblasts. Another important question that needs to be addressed in future is which factor/s released by the tumor cells is/are the driver/s of *THBS4* expression in fibroblasts. However, much points to growth factors, such as transforming growth factor β (TGF-β), platelet-derived growth factor (PDGF) or fibroblast growth factor 2 (FGF2), as the promoters of this effect. All of them are known key regulators of fibroblast activation and are found to be overexpressed in many tumor entities (Kalluri and Zeisberg 2006). In line with that assumption, the transcripts of most of these growth factors and/or their receptors were found to be consistently overexpressed in the diffuse-type tumors of this study compared to intestinal ones (Tab. 16). However, other potential inducers of THBS4 expression in fibroblasts may include cytokines, reactive oxygen species, hypoxia and mechanical stress, which are also usually present in tumor tissue (Orend and Chiquet-Ehrismann 2006).

Table 16: Regulation of selected growth factors and their receptors in human diffuse-type gastric adenocarcinomas.

Gene transcripts were examined for differential expression in a microarray data set comparing diffuse and intestinal-type gastric tumors. Welch-tests with Benjamini and Hochberg FDR (with respect to all probe sets on the array) were performed to evaluate significance in differential expression. Significant genes are framed in red. If genes were covered by more than one probe set, the one showing the strongest significance is presented.

FDR – False Discovery Rate; corr. – corrected; FC – fold change; reg. – regulation; ID – identifier

Gene symbol	Gene title	Corr. *p*-value	FC	Reg. in diff.	Probe set ID	UniGene ID
FGF2	Fibroblast growth factor 2	2.38E-05	4.3	up	204422_s_at	Hs.284244
FGFR1	Fibroblast growth factor receptor 1	1.08E-04	2.8	up	226705_at	Hs.264887
FGFR2	Fibroblast growth factor receptor 2	7.93E-02	1.6	down	208229_at	Hs.533683
FGFR3	Fibroblast growth factor receptor 3	3.04E-01	1.6	down	204379_s_at	Hs.1420
FGFR4	Fibroblast growth factor receptor 4	8.72E-02	1.5	down	1554962_a_at	Hs.533683
PDGFA	Platelet-derived growth factor *alpha* polypeptide	4.97E-01	1.4	up	216867_s_at	Hs.535898
PDGFB	Platelet-derived growth factor *beta* polypeptide	2.18E-01	1.3	up	204200_s_at	Hs.1976
PDGFC	Platelet-derived growth factor C	2.63E-01	1.5	up	218718_at	Hs.570855
PDGFD	Platelet-derived growth factor D	3.58E-01	1.3	up	222860_s_at	Hs.352298
PDGFRA	Platelet-derived growth factor receptor, *alpha* polypeptide	3.80E-03	2.1	up	203131_at	Hs.74615
PDGFRB	Platelet-derived growth factor receptor, *beta* polypeptide	1.41E-05	2.7	up	202273_at	Hs.509067
PDGFRL	Platelet-derived growth factor receptor-like	1.02E-03	5.0	up	205226_at	Hs.458573
TGFB1	Transforming growth factor, *beta* 1	1.17E-01	1.6	up	203085_s_at	Hs.645227
TGFB1I1	Transforming growth factor *beta* 1 induced transcript 1	6.02E-04	2.8	up	209651_at	Hs.513530
TGFB2	Transforming growth factor, *beta* 2	2.66E-02	1.9	up	228121_at	Hs.133379
TGFB3	Transforming growth factor, *beta* 3	4.68E-05	3.5	up	209747_at	Hs.592317
TGFBR1	Transforming growth factor, *beta* receptor 1	1.31E-02	1.6	up	224793_s_at	Hs.494622
TGFBR2	Transforming growth factor, *beta* receptor 2	8.35E-03	1.6	up	208944_at	Hs.82028
TGFBR3	Transforming growth factor, *beta* receptor 3	8.83E-05	3.3	up	226625_at	Hs.482390

To this date, THBS4 has been a molecule literally unknown to the scenario of diffuse-type gastric adenocarcinomas and gastric cancer in general. Except for the above mentioned meta-analysis of three independent gastric cancer microarray data sets, which had assigned a putative overexpression of the *THBS4* gene to the diffuse type (Kim *et al*. 2007b), no publication record exists for this molecule in the context of gastric cancer. However, *THBS4* just emerges as a member of a list of differentially expressed genes without any further validation or examination on protein level in this meta-study. Thus, the here presented work is the first to de-

scribe and characterize THBS4 expression in gastric cancer and to limit its expression to diffuse-type adenocarcinomas. It unraveled that in these carcinomas THBS4 is expressed and secreted by cancer-associated fibroblasts upon activation by tumor cells and is vastly deposited within the extracellular matrix of the tumor stroma. Besides THBS4, many other extracellular matrix proteins are overexpressed and incidentally accumulated in the stroma of diffuse-type gastric adenocarcinomas, according to this and other studies (refer to Tab. 17 appendix page 165 ff.). This phenomenon of increased deposition of extracellular matrix is known as desmoplasia and is frequently accompanying many invasive carcinoma entities. Desmoplastic lesions usually contain increased amounts of collagens, fibronectins and proteoglycans (Kalluri and Zeisberg 2006). In conclusion, THBS4 represents a novel highly abundant constituent of the activated desmoplastic stroma of diffuse-type gastric adenocarcinomas.

However, the function of THBS4 within the reactive stroma of these tumors remains to be elusive. Whether it has a function for tumor cell behavior at all, and if yes, which one/s needs to be investigated in future. Nonetheless, a putative function on tumor cell characteristics can be hypothesized for this protein. Many previous studies were able to assign functions regarding progression, invasion and metastasization of carcinomas to many other extracellular matrix molecules. For example, tenascin C, an extracellular matrix molecule that has been studied in great detail, is known to increase proliferation, angiogenesis, invasion and metastasization of many carcinoma entities (Orend and Chiquet-Ehrismann 2006). In addition, collagen 1A1 and 1A2 are members of a recently identified metastasis gene signature that had been trained on a compendium of different solid tumor entities (Ramaswamy *et al*. 2003). Within the signature, both molecules were associated with metastasis formation when upregulated. However, some constituents of the extracellular matrix have also been shown to possess adverse effects, such as the inhibition of tumor progression (Barsky and Gopalakrishna 1987). In addition to modifying the migratory behavior of tumor cells, extracellular matrix proteins also affect various regulatory pathways by modulating release and availabil-

ity of growth factors and by direct interactions with other cells through integrins (Kalluri and Zeisberg 2006).

In non-tumor environments, limited information about the physiological functions of THBS4 is available, indeed. For example, it was found to be expressed by a wide range of neurons in the adult and developing nervous system and to promote neurite outgrowth (Arber and Caroni 1995). Furthermore, proliferation of erythroid cells, hematopoietic precursors (CD34-positive cells), skin fibroblasts and kidney epithelial cells was identified to be stimulated by THBS4 (Congote *et al*. 2004, Sadvakassova *et al*. 2009). However, also anti-proliferatory effects of the protein were observed, namely on endothelial cells (Congote *et al*. 2004). Other proposed functions include the support of myoblast adhesion (Adams and Lawler 1994) and the interaction with other extracellular matrix proteins, such as certain collagens, laminin α 1, fibronectin and matrilin 2 (Narouz-Ott *et al*. 2000). Given these information, hypothetical functions of THBS4 within cancer scenarios could include stimulation of proliferation of tumor cells and/or of other cell types and facilitation of migratory behavior and, potentially also, of invasive potential of tumor cells. Such hypothetical functions could either be implemented in a direct manner through binding of THBS4 to a specific receptor followed by activation of the downstream signaling cascade or indirectly via interactions and/or cross reactions with other extracellular matrix proteins. If these hypothetical functions hold true, THBS4 could represent one of the key molecules enabling the vast invasiveness of diffuse-type gastric adenocarcinomas and thus a potentially promising novel target for anti-cancer therapy.

Besides diffuse-type gastric adenocarcinomas, also lobular carcinomas of the breast, mesoblastic nephroma, Wilms tumors and grade I pilocytic astrocytomas were shown to exhibit overexpression of THBS4, at least on transcriptional level (Korkola *et al*. 2003, Rorive *et al*. 2006, Sugimura *et al*. 2004, Turashvili *et al*. 2005). Via the ONCOMINE database, a compendium of cancer transcriptome profiles with web-based data-mining platform (Rhodes *et al*. 2004), additional

significant mRNA overexpression could be identified for prostate carcinomas. This overexpression in multiple additional tumor entities suggests that THBS4 might be of more global malignant relevance rather than being restricted to diffuse-type cancers of the stomach, solely. Of above-mentioned cancer entities, overexpression in lobular carcinomas of the breast is evidenced the best so far. Two microarray-based expression profiling studies comparing lobular to ductal carcinomas of the breast found the *THBS4* transcript to be significantly higher present in the lobular type (Korkola *et al*. 2003, Turashvili *et al*. 2005). In the ONCOMINE database, additional 3 microarray studies showing overexpression of *THBS4* in lobular carcinomas compared to ductal ones could be identified (Lu *et al*. 2008, Ma *et al*. 2004, Sorlie *et al*. 2003). Given the fact that lobular carcinomas of the breast grow in a diffuse manner as well, this suggests that THBS4 upregulation and accumulation might be events frequently associated with diffusely growing cancers, especially.

4.3 Prognostic gene signatures and marker genes for N+ gastric adenocarcinomas

Involvement of regional lymph nodes is a condition commonly accompanying diagnosis of gastric cancer and is currently considered the most important parameter for assessment of prognosis and choice of therapy. Patients with positive lymph nodes (N+) generally possess poor outcomes, whereas patients without lymphatic dissemination clearly own more favorable prognosis. However, not all lymph node-positive patients exhibit poor prognosis. A small subgroup of these patients survives 5 years and longer (Kim *et al*. 2007a). Currently, there are no clinicopathological parameters available that clearly predict these differences in survival time among lymph node-positive patients (Cheong *et al*. 2006, Kim *et al*. 2007a). Thus, this study had aimed at the stratification of these patients subgroups

on the molecular level and at the identification of putative marker genes or gene signatures that allow a more accurate assessment of prognosis.

Such molecular markers or signatures for prognosis may lead the way to more personalized prognosis assessment and thus to more tailored therapeutic approaches in future. As a result, the amounts of patients being over or undertreated could be reduced, which clearly bears benefits for patients but also for the national health systems.

The identification of prognostic genes and signatures was performed on a very stringently selected patient cohort. Additional to the basic selection criterion of diagnosis with regional lymph node-positive gastric adenocarcinoma (N+), all patients had been treated with curative intention without neoadjuvant or adjuvant chemotherapy, had undergone complete resection with no residual tumor left (R0), had no distant metastases by the time of diagnoses and surgery (M0) and no diagnosed secondary malignancy. Progression-based criteria, which needed to be fulfilled, included "absence of postoperative death" and "appearance of first recurrence of disease at least 2 months after surgery". Thus, the possibility that obtained results are confounded by systematic errors caused by such clinical parameters is minimized in this study.

Patients were split into two discrete cohorts according to histological type of tumor and the identification of putative prognostic genes and signatures was performed separately for the cohorts. This is a novelty, since all previous gene expression profiling studies that had aimed at the identification of prognostic markers for gastric cancer had not been conducted on histologically stratified patient cohorts (Chen *et al*. 2005, Chen *et al*. 2003, Lee *et al*. 2010, Leung *et al*. 2002, Leung *et al*. 2004, Takeno *et al*. 2010).

By performance of statistical significance testing comparing each gene's average expression in the poor prognosis group to its average in the good prognosis group, genes correlating with prognosis could be extracted from the microarray data for either histological cohort. However, none of these genes passed any multiple test-

ing procedures controlling the number of false-positives. The *p*-value cutoff of $p<0.001$, which was set for the selection procedure, yields ~54 false-positive probe sets when introducing all probe sets on the array into the test. Hence, the numbers of probe sets identified as correlating with prognosis, namely 36 for intestinal and 18 for diffuse-type patients, are in the false-positive range of this test system. Therefore, the genes and the whole signatures themselves cannot be discerned from false-positives at this time point. However, the statistical significance of individual genes was stronger for the intestinal-type patient cohort, with *p*-values ranging up to 10^{-7}, whereas *p*-values for the diffuse cohort yielded values of 10^{-5} at best. Furthermore, also more genes putatively correlating with prognosis could be identified for the intestinal cohort than for the diffuse one. These facts indicate that prognosis assessment based on the gene expression profile of the primary tumor might be more useful and suitable for intestinal-type N+ gastric adenocarcinoma patients than for diffuse-type ones. The reasons for this might be of true biological nature, but also of systematic origin. Concerning biology, it is well conceivable that diffuse tumors do not contain molecularly diverse subgroups whose differences are reflected on prognosis level, since they represent a very aggressive phenotype with poor prognosis inherently. For intestinal tumors, in contrast, the existence of subgroups with distinct expression profiles that lead to differences in patients' prognosis is more reasonable. Regarding systematic errors confounding the results, one example could be a "hidden" difference in resection status. Complete resection is inherently more difficult to achieve in diffuse-type tumors than in intestinal ones, because single tumor cells need to be detected by the pathologist at the tumor margin, which can be exceedingly challenging. In order to account for this problem, larger tumor margins are resected for diffuse-type tumors than for intestinal ones. However, whether this procedure fully compensates for the risk of residual tumor in these tumors seems to be questionable. In principle, the proportion of mistaken complete resections might be higher in diffuse tumors than in intestinal ones. Certainly, the reason for recurrence and poor prognosis of these non-complete resected tumors is not of molecular nature,

but driven by the residual tumor cells. Another systematic factor contributing to the observed differences in significance could be the amount of "contaminating" non-tumor cells, which is often much higher in diffuse than in intestinal tumors. Such non-tumor cells include, for example, fibroblasts, which are strongly recruited and activated by diffuse tumors and cause the desmoplastic phenotype that is often associated with it. As a consequence the measured expression profile of diffuse tumors might be tampered with the expression profile of non-tumor cells in a more prominent manner as for intestinal-type tumors.

From the panel of genes possessing putative prognostic value, certain candidates were chosen for validation by means of quantitative real-time PCR. The candidates comprised the growth associated protein 43 (*GAP43*) and the ephrin receptor A4 (*EPHA4*) for the diffuse-type patient cohort and ras-related associated with diabetes (*RRAD*), homeobox C10 (*HOXC10*), RAN binding protein 17 (*RANBP17*) and the folate receptor 1 (*FOLR1*) for the intestinal-type one. According to quantitative real-time PCR data, a clear value for distinguishing patients with poor prognosis from long-term disease-free survivors could be assigned to all candidates selected for the intestinal cohort. The strongest power for separation of the groups could be identified for *RANBP17*, indeed. Multivariate analyses confirmed that each gene's prognostic value is independent of T-stage and N-stage in this patient cohort. In contrast, none of the candidates selected for the diffuse-type patients showed any true capability to discriminate the prognostic groups, respectively. Interestingly, *RRAD,* one of the candidates originally selected for the intestinal cohort, exhibited the ability to discern the prognostic groups of the diffuse cohort as well. Thus, *RRAD* might be the only one of the selected candidate genes which offers prognostic value for both histological types.

The quantitative real-time PCR validation confirmed the trend, which had already been encountered during microarray data analysis, that stronger separation efficiencies regarding prognosis are present for the candidates selected for the intestinal patients. This strengthens once more the idea that prognosis assessment based

on the gene expression profile of primary tumors is potentially more appropriate for intestinal-type N+ gastric adenocarcinomas than for diffuse-type ones.

Consistent for both histological cohorts of this study is that none of the selected candidates was able to allow an overall and exclusive separation of patients into groups of different prognosis. Every candidate was upregulated in subpopulations of patients possessing poor prognosis, only. This indicates that the molecular reasons for recurrence of disease in these patients is rather heterogeneous than homogeneous. For example, it seems to be unlikely that the activation of the same individual pathway or signaling cascade causes recurrence in all patients. More probable is the scenario that different molecular mechanisms triggering recurrence exist. This idea is further strengthened by the finding that no firm biological connection could be encountered for the members of the gene signatures. The signatures consisted of a more or less mixed population of genes in terms of biology and function. The phenomenon that prognosis assessment on the basis of one individual molecular marker is more or less not feasible is not new. It had been encountered by basically all previous studies that were also seeking for prognostic marker or classifier genes in gastric cancer and other cancer entities. Therefore, it had been proposed that a combination of genes, a so-called gene signature, is in principle needed for clear prediction of prognosis. The here identified gene signatures were certainly more precise in distinguishing the different prognostic groups of this study than single members of it alone, however, nothing about any global prognostic value of the signatures can be states at this time point.

Future studies need to reveal whether the here identified gene signatures and/or single markers are purely cohort-specific or are able to predict prognosis in a broader range of N+ gastric adenocarcinoma patients. The possibility that the identified signatures and markers are of cohort-specific nature is especially conceivable considering the rather small patient numbers on whose basis they had been identified. The good prognosis (non-recurrent) patient groups of both cohorts were especially small. However, this is due to the biological nature of N+ patients,

which in majority exhibit poor outcomes. The next fundamental step would be to validate the prognostic signatures and markers in additional patient cohorts, so-called independent test sets.

In principle, a postulated prognostic marker gene or gene signature needs to be validated in multiple independent patient cohorts in order to fully establish it as a prognostic marker or classifier. Only if the prognostic value can be confirmed by such further studies, a gene or gene signature might enter into clinical routine as an established marker.

RANBP17, whose transcriptional expression showed by far the strongest capability to indicate differences in prognosis of the intestinal-type cohort of this study, is a member of the importin-*beta* superfamily. Importin-*beta* proteins are subunits of the importin nuclear transport receptors, which import proteins (with a nuclear localization signal) through the nuclear pore complexes from the cytoplasm to the nucleus. Most importins are heterodimers of importin-*beta* and importin-*alpha* proteins and all of them require interaction with the small GTPase RAN to operate (Terry *et al*. 2007). Proteins that usually need to be imported to the nucleus after translation in the cytoplasm comprise histones, ribosomal proteins, signalling proteins (*e.g.* transcription factors) and splicing factors for mRNA processing; all of which fulfil certain nuclear functions. Thus, alterations and defects in this subcellular transport mechanism may lead to changes in DNA packaging, ribosome assembly, transcriptional activity, mRNA splicing *etc*. and consequently to gene deregulation. Malfunctions, alterations or impairment of cytoplasmic-nuclear-transport have been detected in many different types of cancer cells, indicating that proper function of this process is important for sustaining normal non-transformed cell function.

Dysexpression of importins and their subunits may be one of the reasons causing malfunctions of cytoplasmic-nuclear-transport and associated nuclear processes. Overexpression of RANBP17, as observed in this study, may lead to increased import of certain transcription factors and/or other transcription-relevant proteins

and subsequently to increased transcription of respective target genes. However, so far nothing is known about which particular proteins are transported by RANBP17. For other members of the RANBP family, such interaction partners are known already and comprise, for example, β-catenin and TCF-4, the nuclear effectors of Wnt signaling (Hendriksen *et al*. 2005, Shitashige *et al*. 2008).

To this date, deregulation of *RANBP17* has never been encountered in gastric cancer or any other cancer of the gastrointestinal tract. Thus, this study is the first to identify this protein to be putatively deregulated and differentially expressed in this disease and to potentially be of prognostic value for N+ patients. Regarding other cancer entities, reports exists about deregulation of *RANBP17* in certain types of leukemia (Bernard *et al*. 2001). Via the OCOMINE database (Rhodes *et al*. 2004), the mRNA expression of *RANBP17* could be examined in a broad panel of cancer entities. This screen elucidated that *RANBP17* is transcriptionally deregulated in various cancer entities (Fig. 36). However, no gastric cancer data set featuring the *RANBP17* transcript was available in the database. No significant correlation with prognosis or clinical outcome could be identified for *RANBP17* expression in the data sets available.

The function of RANBP17 within cancer scenarios is completely elusive to this date.

To my best knowledge, this work is the first to gene expression-profile exclusively and specifically node-positive gastric adenocarcinoma patients and to identify candidate genes and signatures correlating with clinical outcome and prognosis of these patients. Thus, it represent an initial step in elucidating molecular mechanisms that are potentially involved in the progression of this “advanced” stage of disease and in the establishment of prognostic signatures, markers and/or classifiers for these patients in future.

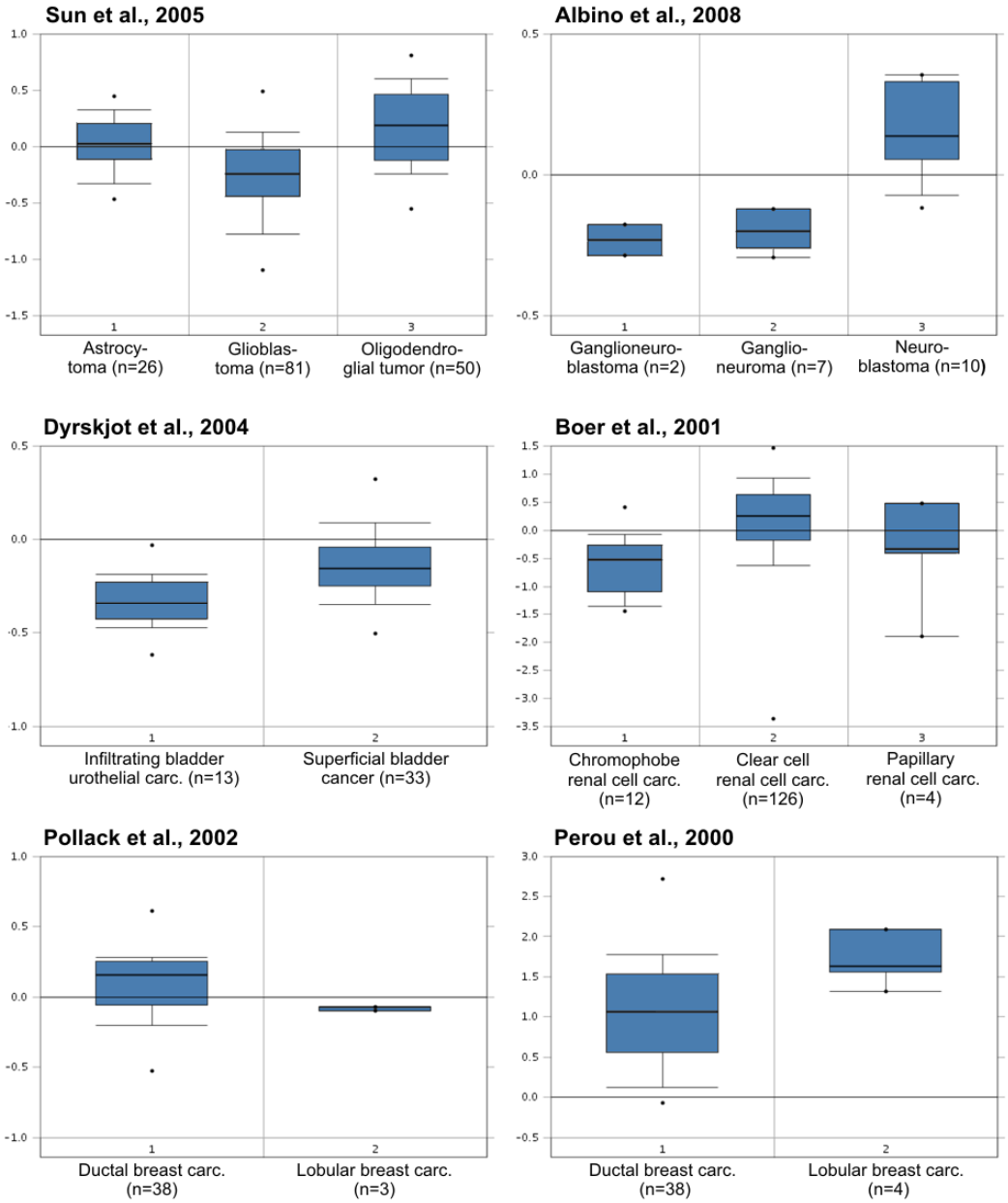


Figure 36: *RANBP17* mRNA expression in different cancer entities and their subtypes.

The ONCOMINE database (Rhodes *et al*. 2004) was searched for studies possessing significant (p<0.01) differential *RANBP17* mRNA expression between tumor subtypes. Examples of such studies are depicted. Box and whisker plot images illustrating the distribution of *RANBP17* mRNA expression within different subgroups were extracted from the database.

carc. – carcinoma; n – number

4.4 BMP and activin membrane-bound inhibitor homolog as a prognostic gene for N+ gastric adenocarcinomas

BAMBI is a pseudoreceptor that is related to the TGF-*β* receptor type I (TGFBR1), but lacks an intracellular kinase domain that is important for activating intracellular SMADs (Onichtchouk *et al*. 1999). Thus, BAMBI acts as a negative regulator of TGF-*β* (TGFB1) signaling. The aberrant elevation of BAMBI, which is found in most colorectal cancers, is a result of expression induction via Wnt signaling through *β*-catenin (CTNNB1) activation. Hence, Wnt signaling transcriptionally activates BAMBI as a mechanism to inhibit TGF-*β* signaling. However, BAMBI can also be transcriptionally induced by TGF-*β*, through SMAD-binding elements in the *BAMBI* promoter (Sekiya *et al*. 2004), probably as a countermeasure to negatively regulate its own pathway. Besides mediation of the TGF-*β* pathway, also mediation of the Wnt pathway, but in a promotional manner as positive feedback regulation was evidenced for BAMBI. Thereby, increased Wnt/*β*-catenin transcriptional activity leads to increased expression of growth promoting targets, such as c-myc (MYC), cyclin D1 (CCND1) and D2 (CCND2). In summary, BAMBI can be induced by Wnt and TGF-*β*-SMAD signaling to positively regulate Wnt signaling and negatively regulate TGF-*β* signaling. The overall effect of BAMBI would be to increase cellular growth through promotion of Wnt proliferative signaling and inhibition of TGF-*β* growth suppressive signaling (Fig. 37).

In colorectal cancer, patients featuring “high” levels of *BAMBI* mRNA expression are prone to a significantly shortened metastasis-free survival compared to “low” expressing patients (Fritzmann *et al*. 2009). However, BAMBI not just serves as a marker for postoperative metastases formation, but actually plays an active role in the metastasization process (Fritzmann *et al*. 2009, Togo *et al*. 2008). Thus, it represents an attractive molecule for personalized prognosis prediction and a promising target for therapeutic interventions.

So far nothing is known about BAMBI expression and its prognostic value in gastric cancer. Given these objectives, this study had aimed at a preliminary evaluation of BAMBI expression in gastric adenocarcinomas and its prognostic meaning for N+ patients.

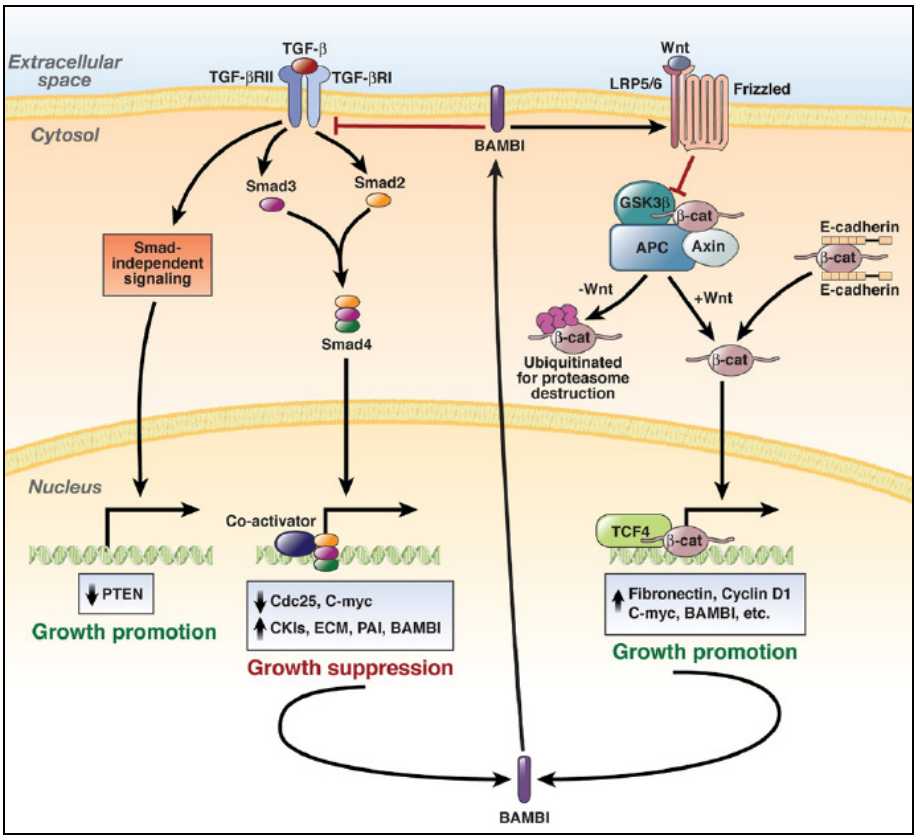


Figure 37: Schematic overview of Wnt and TGF-*β* signaling pathways, and the regulatory role of BAMBI (Carethers 2009).

Quantitative real-time PCR analysis revealed that *BAMBI* mRNA is in general much higher expressed in intestinal than in diffuse-type gastric adenocarcinomas. Only single diffuse-type tumors possess comparable amounts as present in intestinal type. Regarding prognosis, patients featuring “high” or “low” *BAMBI* expression showed significant differences in metastasis-free survival, with longer survival times present in the group with “low” expression. This difference could be

observed for both histological patient cohorts. However, stronger significance was encountered in the diffuse-type patients. Hence, this study is the first to show that *BAMBI* mRNA expression bears prognostic value for N+ gastric adenocarcinomas and serves as a marker for metastasis-free survival in these patients. However, larger and independent patient cohorts are needed to confirm this data.

In situ hybridization experiments localized *BAMBI* mRNA expression to the tumor cell population of both histological types. *BAMBI* was expressed by basically all cells of the discrete tumor in intestinal-type tumors, whereas only single tumor cells showed expression in the diffuse type. The intensity of stained cells was rather equal in both types. Thus, both histological types do not differ in the extent of *BAMBI* expression on single cell-level, but in the fraction of tumor cells owning expression. Why only certain tumor cells of the diffuse type exhibit *BAMBI* expression needs to be elucidated in future. However, due to the difference in expression pattern between the two types, it could be postulated that BAMBI might be regulated differently in each type and, along with that, play different roles in initiation and/or progression of each type. Since healthy gastric mucosa counterparts did not feature any *BAMBI* expression, *BAMBI* expression is a gastric adenocarcinoma-specific event.

This study is the first to analyze *BAMBI* mRNA expression in gastric adenocarcinomas and to elucidate that it is of prognostic relevance for N+ patients. Thus, this study provides a first indication that the metastasis-promoting features of BAMBI can be expanded to further cancer entities, at least to certain cancers of the stomach.

Bibliography

Adachi Y, Shiraishi N, Suematsu T, Shiromizu A, Yamaguchi K, Kitano S (2000a). Most important lymph node information in gastric cancer: multivariate prognostic study. *Ann Surg Oncol* **7:** 503-507.

Adachi Y, Yasuda K, Inomata M, Sato K, Shiraishi N, Kitano S (2000b). Pathology and prognosis of gastric carcinoma: well versus poorly differentiated type. *Cancer* **89:** 1418-1424.

Adams JC, Lawler J (1994). Cell-type specific adhesive interactions of skeletal myoblasts with thrombospondin-1. *Mol Biol Cell* **5:** 423-437.

Ajisaka H, Miwa K (2003). Micrometastases in sentinel nodes of gastric cancer. *Br J Cancer* **89:** 676-680.

Allard JE, Risinger JI, Morrison C, Young G, Rose GS, Fowler J *et al.* (2007). Overexpression of folate binding protein is associated with shortened progression-free survival in uterine adenocarcinomas. *Gynecol Oncol* **107:** 52-57.

American Cancer Society (2007). Global Cancer Facts & Figures 2007.

Arber S, Caroni P (1995). Thrombospondin-4, an extracellular matrix protein expressed in the developing and adult nervous system promotes neurite outgrowth. *J Cell Biol* **131:** 1083-1094.

Arias AM (2001). Epithelial mesenchymal interactions in cancer and development. *Cell* **105:** 425-431.

Barsky SH, Gopalakrishna R (1987). Increased invasion and spontaneous metastasis of BL6 melanoma with inhibition of the desmoplastic response in C57 BL/6 mice. *Cancer Res* **47:** 1663-1667.

Benjamini Y, Hochberg Y (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society Series B (Methodological)* **57:** 289-300.

Bernard OA, Busson-LeConiat M, Ballerini P, Mauchauffe M, Della Valle V, Monni R *et al.* (2001). A new recurrent and specific cryptic translocation, t(5;14)(q35;q32), is associated with expression of the Hox11L2 gene in T acute lymphoblastic leukemia. *Leukemia* **15:** 1495-1504.

Bluthgen N, Brand K, Cajavec B, Swat M, Herzel H, Beule D (2005). Biological profiling of gene groups utilizing Gene Ontology. *Genome Inform* **16:** 106-115.

Bolstad BM, Irizarry RA, Astrand M, Speed TP (2003). A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics* **19:** 185-193.

Bonferroni CE (1935). Il calcolo delle assicurazioni su gruppi di teste. *"Studi in Onore del Professore Salvatore Ortu Carboni."* Rome. pp 13-60.

Bonferroni CE (1936). Teoria statistica delle classi e calcolo delle probabilita. *Pubblicazioni del R Instituto Superiore de Scienze Economiche e Commerciali de Firenze***:** 3-62.

Boussioutas A, Li H, Liu J, Waring P, Lade S, Holloway AJ *et al.* (2003). Distinctive patterns of gene expression in premalignant gastric mucosa and gastric cancer. *Cancer Res* **63:** 2569-2577.

Carethers JM (2009). Intersection of transforming growth factor-beta and Wnt signaling pathways in colorectal cancer and metastasis. *Gastroenterology* **137:** 33-36.

Carneiro F (1997). Classification of gastric carcinomas. *Current Diagnostic Pathology* **4:** 51-59.

Chan AO, Luk JM, Hui WM, Lam SK (1999). Molecular biology of gastric carcinoma: from laboratory to bedside. *J Gastroenterol Hepatol* **14:** 1150-1160.

Chen CN, Lin JJ, Chen JJ, Lee PH, Yang CY, Kuo ML *et al.* (2005). Gene expression profile predicts patient survival of gastric cancer after surgical resection. *J Clin Oncol* **23:** 7286-7295.

Chen X, Leung SY, Yuen ST, Chu KM, Ji J, Li R *et al.* (2003). Variation in gene expression patterns in human gastric cancers. *Mol Biol Cell* **14:** 3208-3215.

Cheong JH, Hyung WJ, Shen JG, Song C, Kim J, Choi SH *et al.* (2006). The N ratio predicts recurrence and poor prognosis in patients with node-positive early gastric cancer. *Ann Surg Oncol* **13:** 377-385.

Congote LF, Difalco MR, Gibbs BF (2004). The C-terminal peptide of thrombospondin-4 stimulates erythroid cell proliferation. *Biochem Biophys Res Commun* **324:** 673-678.

Correa P (1988). A human model of gastric carcinogenesis. *Cancer Res* **48:** 3554-3560.

Correa P, Shiao YH (1994). Phenotypic and genotypic events in gastric carcinogenesis. *Cancer Res* **54:** 1941s-1943s.

Davessar K, Pezzullo JC, Kessimian N, Hale JH, Jauregui HO (1990). Gastric adenocarcinoma: prognostic significance of several pathologic parameters and histologic classifications. *Hum Pathol* **21:** 325-332.

Day D, Morson B, Jass J, Price A, N. S, Williams G *et al.* (2003). *Morson and Dawson's Gastrointestinal Pathology*, 4 edn. Wiley-Blackwell (Blackwell Publishing).

Dicken BJ, Bigam DL, Cass C, Mackey JR, Joy AA, Hamilton SM (2005). Gastric adenocarcinoma: review and considerations for future directions. *Ann Surg* **241:** 27-39.

Erenoglu C, Akin ML, Uluutku H, Tezcan L, Yildirim S, Batkin A (2000). Angiogenesis predicts poor prognosis in gastric carcinoma. *Dig Surg* **17:** 581-586.

Evans SJ, Watson SJ, Akil H (2003). Evaluation of Sensitivity, Performance and Reproducibility of Microarray Technology in Neuronal Tissue. *Integrative and Comparative Biology* **43:** 780-785.

Ferlay J, Bray F, Pisani P (2004). GLOBOCAN 2002. Cancer Incidence, Mortality and Prevalence Worldwide. *IARC Cancer Base No 5 Version 20 Lyon: International Agency for Research on Cancer*.

Fritzmann J, Morkel M, Besser D, Budczies J, Kosel F, Brembeck FH *et al.* (2009). A colorectal cancer expression profile that includes transforming growth factor beta inhibitor BAMBI predicts metastatic potential. *Gastroenterology* **137:** 165-175.

Gabbiani G, Ryan GB, Majne G (1971). Presence of modified fibroblasts in granulation tissue and their possible role in wound contraction. *Experientia* **27:** 549-550.

Gallagher SR, Desjardins PR (2006). Quantitation of DNA and RNA with absorption and fluorescence spectroscopy. *Curr Protoc Mol Biol* Appendix 3: Appendix 3D.

Ganten D, Ruckpaul K (1998). *Tumorerkrankungen*. Springer: Berlin.

Goseki N, Takizawa T, Koike M (1992). Differences in the mode of the extension of gastric cancer classified by histological type: new histological classification of gastric carcinoma. *Gut* **33:** 606-612.

Guo X, Oshima H, Kitmura T, Taketo MM, Oshima M (2008). Stromal fibroblasts activated by tumor cells promote angiogenesis in mouse gastric cancer. *J Biol Chem* **283:** 19864-19871.

Hamilton SR, Aaltonen LA (eds) (2000) *Pathology and Genetics. Tumours of the Digestive System*. IARC Press: Lyon.

Hasegawa S, Furukawa Y, Li M, Satoh S, Kato T, Watanabe T *et al.* (2002). Genome-wide analysis of gene expression in intestinal-type gastric cancers using a complementary DNA microarray representing 23,040 genes. *Cancer Res* **62:** 7012-7017.

Hattori T (1986). Development of adenocarcinomas in the stomach. *Cancer* **57:** 1528-1534.

Hendriksen J, Fagotto F, van der Velde H, van Schie M, Noordermeer J, Fornerod M (2005). RanBP3 enhances nuclear export of active (beta)-catenin independently of CRM1. *J Cell Biol* **171:** 785-797.

Henson DE, Dittus C, Younes M, Nguyen H, Albores-Saavedra J (2004). Differential trends in the intestinal and diffuse types of gastric carcinoma in the United States, 1973-2000: increase in the signet ring cell type. *Arch Pathol Lab Med* **128:** 765-770.

Hippo Y, Taniguchi H, Tsutsumi S, Machida N, Chong JM, Fukayama M *et al.* (2002). Global gene expression analysis of gastric cancer by oligonucleotide microarrays. *Cancer Res* **62:** 233-240.

Hohenberger P, Gretschel S (2003). Gastric cancer. *Lancet* **362:** 305-315.

Hundahl SA, Phillips JL, Menck HR (2000). The National Cancer Data Base Report on poor survival of U.S. gastric carcinoma patients treated with gastrectomy: Fifth Edition American Joint Committee on Cancer staging, proximal disease, and the "different disease" hypothesis. *Cancer* **88:** 921-932.

IARC (2008). International Agency for Research on Cancer: World Cancer Report 2008.

Irizarry RA, Bolstad BM, Collin F, Cope LM, Hobbs B, Speed TP (2003a). Summaries of Affymetrix GeneChip probe level data. *Nucleic Acids Res* **31:** e15.

Irizarry RA, Hobbs B, Collin F, Beazer-Barclay YD, Antonellis KJ, Scherf U *et al.* (2003b). Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics* **4:** 249-264.

Iwakiri S, Sonobe M, Nagai S, Hirata T, Wada H, Miyahara R (2008). Expression status of folate receptor alpha is significantly correlated with prognosis in non-small-cell lung cancers. *Ann Surg Oncol* **15:** 889-899.

Jansen EP, Boot H, Verheij M, van de Velde CJ (2005). Optimal locoregional treatment in gastric cancer. *J Clin Oncol* **23:** 4509-4517.

Jinawath N, Furukawa Y, Hasegawa S, Li M, Tsunoda T, Satoh S *et al.* (2004). Comparison of gene-expression profiles between diffuse- and intestinal-type gastric cancers using a genome-wide cDNA microarray. *Oncogene* **23:** 6830-6844.

Juttner S, Wissmann C, Jons T, Vieth M, Hertel J, Gretschel S *et al.* (2006). Vascular endothelial growth factor-D and its receptor VEGFR-3: two novel independent prognostic markers in gastric adenocarcinoma. *J Clin Oncol* **24:** 228-240.

Kalluri R, Zeisberg M (2006). Fibroblasts in cancer. *Nat Rev Cancer* **6:** 392-401.

Kaplan EL, Meier P (1958). Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association* **53:** 457-481.

Kazerounian S, Yee KO, Lawler J (2008). Thrombospondins in cancer. *Cell Mol Life Sci* **65:** 700-712.

Kim DY, Seo KW, Joo JK, Park YK, Ryu SY, Kim HR *et al.* (2006). Prognostic factors in patients with node-negative gastric carcinoma: a comparison with node-positive gastric carcinoma. *World J Gastroenterol* **12:** 1182-1186.

Kim DY, Joo JK, Park YK, Ryu SY, Kim YJ, Kim SK (2007a). Predictors of long-term survival in node-positive gastric carcinoma patients with curative resection. *Langenbecks Arch Surg* **392:** 131-134.

Kim HJ, Kim AY, Oh ST, Kim JS, Kim KW, Kim PN *et al.* (2005). Gastric cancer staging at multi-detector row CT gastrography: comparison of transverse and volumetric CT scanning. *Radiology* **236:** 879-885.

Kim MC, Kim HH, Jung GJ, Lee JH, Choi SR, Kang DY *et al.* (2004). Lymphatic mapping and sentinel node biopsy using 99mTc tin colloid in gastric cancer. *Ann Surg* **239:** 383-387.

Kim SY, Kim JH, Lee HS, Noh SM, Song KS, Cho JS *et al.* (2007b). Meta- and gene set analysis of stomach cancer gene expression data. *Mol Cells* **24:** 200-209.

Koch WH (2004). Technology platforms for pharmacogenomic diagnostic assays. *Nat Rev Drug Discov* **3:** 749-761.

Korkola JE, DeVries S, Fridlyand J, Hwang ES, Estep AL, Chen YY *et al.* (2003). Differentiation of lobular versus ductal breast carcinomas by expression microarray analysis. *Cancer Res* **63:** 7167-7175.

Lackie J (2007). *The Dictionary of Cell & Molecular Biology*, 4th edition edn. Academic Press

Lauren P (1965). The Two Histological Main Types of Gastric Carcinoma: Diffuse and So-Called Intestinal-Type Carcinoma. an Attempt at a Histo-Clinical Classification. *Acta Pathol Microbiol Scand* **64:** 31-49.

Lauren PA, Nevalainen TJ (1993). Epidemiology of intestinal and diffuse types of gastric carcinoma. A time-trend study in Finland with comparison between studies from high- and low-risk areas. *Cancer* **71:** 2926-2933.

Lawler J, McHenry K, Duquette M, Derick L (1995). Characterization of human thrombospondin-4. *J Biol Chem* **270:** 2809-2814.

Lee CC, Wu CW, Lo SS, Chen JH, Li AF, Hsieh MC *et al.* (2007). Survival predictors in patients with node-negative gastric carcinoma. *J Gastroenterol Hepatol* **22:** 1014-1018.

Lee HJ, Nam KT, Park HS, Kim MA, Lafleur BJ, Aburatani H *et al.* (2010). Gene expression profiling of metaplastic lineages identifies CDH17 as a prognostic marker in early stage gastric cancer. *Gastroenterology* **139:** 213-225 e213.

Lee KJ, Inoue M, Otani T, Iwasaki M, Sasazuki S, Tsugane S (2006). Gastric cancer screening and subsequent risk of gastric cancer: a large-scale population-based cohort study, with a 13-year follow-up in Japan. *Int J Cancer* **118:** 2315-2321.

Leung SY, Chen X, Chu KM, Yuen ST, Mathy J, Ji J *et al.* (2002). Phospholipase A2 group IIA expression in gastric adenocarcinoma is associated with prolonged survival and less frequent metastasis. *Proc Natl Acad Sci U S A* **99:** 16203-16208.

Leung SY, Yuen ST, Chu KM, Mathy JA, Li R, Chan AS *et al.* (2004). Expression profiling identifies chemokine (C-C motif) ligand 18 as an independent prognostic indicator in gastric cancer. *Gastroenterology* **127:** 457-469.

Liotta LA, Kohn EC (2001). The microenvironment of the tumour-host interface. *Nature* **411:** 375-379.

Lu X, Lu X, Wang ZC, Iglehart JD, Zhang X, Richardson AL (2008). Predicting features of breast cancer with gene expression patterns. *Breast Cancer Res Treat* **108:** 191-201.

Ma XJ, Wang Z, Ryan PD, Isakoff SJ, Barmettler A, Fuller A *et al.* (2004). A two-gene expression ratio predicts clinical outcome in breast cancer patients treated with tamoxifen. *Cancer Cell* **5:** 607-616.

Macdonald JS, Smalley SR, Benedetti J, Hundahl SA, Estes NC, Stemmermann GN *et al.* (2001). Chemoradiotherapy after surgery compared with surgery alone for adenocarcinoma of the stomach or gastroesophageal junction. *N Engl J Med* **345:** 725-730.

Maehara Y, Kakeji Y, Koga T, Emi Y, Baba H, Akazawa K *et al.* (2002). Therapeutic value of lymph node dissection and the clinical outcome for patients with gastric cancer. *Surgery* **131:** S85-91.

Marchet A, Mocellin S, Belluco C, Ambrosi A, DeMarchi F, Mammano E *et al.* (2007). Gene expression profile of primary gastric cancer: towards the prediction of lymph node status. *Ann Surg Oncol* **14:** 1058-1064.

Maruyama K, Gunven P, Okabayashi K, Sasako M, Kinoshita T (1989). Lymph node metastases of gastric cancer. General pattern in 1931 patients. *Ann Surg* **210:** 596-602.

Matrisian LM, Cunha GR, Mohla S (2001). Epithelial-stromal interactions and tumor progression: meeting summary and future directions. *Cancer Res* **61:** 3844-3846.

McCulloch P, Niita ME, Kazi H, Gama-Rodrigues JJ (2005). Gastrectomy with extended lymphadenectomy for primary treatment of gastric cancer. *Br J Surg* **92:** 5-13.

McLean A (2004). *Imaging in oncology (chapter "Gastric cancer")*, 2 edn, vol. 1. Taylor and Francis.

Meining A, Morgner A, Miehlke S, Bayerdorffer E, Stolte M (2001). Atrophy-metaplasia-dysplasia-carcinoma sequence in the stomach: a reality or merely an hypothesis? *Best Pract Res Clin Gastroenterol* **15:** 983-998.

Metz CE (1978). Basic principles of ROC analysis. *Semin Nucl Med* **8:** 283-298.

Ming SC (1977). Gastric carcinoma. A pathobiological classification. *Cancer* **39:** 2475-2485.

Mori M, Shimada H, Gunji Y, Matsubara H, Hayashi H, Nimura Y *et al.* (2004). S100A11 gene identified by in-house cDNA microarray as an accurate predictor of lymph node metastases of gastric cancer. *Oncol Rep* **11:** 1287-1293.

Msika S, Benhamiche AM, Jouve JL, Rat P, Faivre J (2000). Prognostic factors after curative resection for gastric cancer. A population-based study. *Eur J Cancer* **36:** 390-396.

Mueller MM, Fusenig NE (2004). Friends or foes - bipolar effects of the tumour stroma in cancer. *Nat Rev Cancer* **4:** 839-849.

Mulligan RM (1972). Histogenesis and biologic behavior of gastric carcinoma. *Pathol Annu* **7:** 349-415.

Munoz N, Connelly R (1971). Time trends of intestinal and diffuse types of gastric cancer in the United States. *Int J Cancer* **8:** 158-164.

Munoz N, Franceschi S (1997). Epidemiology of gastric cancer and perspectives for prevention. *Salud Publica Mex* **39:** 318-330.

Nakagawa H, Liyanarachchi S, Davuluri RV, Auer H, Martin EW, Jr., de la Chapelle A *et al.* (2004). Role of cancer-associated stromal fibroblasts in metastatic colon cancer to the liver and their expression profiles. *Oncogene* **23:** 7366-7377.

Narouz-Ott L, Maurer P, Nitsche DP, Smyth N, Paulsson M (2000). Thrombospondin-4 binds specifically to both collagenous and non-collagenous extracellular matrix proteins via its C-terminal domains. *J Biol Chem* **275:** 37110-37117.

National Cancer Institute (2002). Surveillance, epidemiology, and end results: incidence, stomach cancer.

Norsett KG, Laegreid A, Midelfart H, Yadetie F, Erlandsen SE, Falkmer S *et al.* (2004). Gene expression based classification of gastric carcinoma. *Cancer Lett* **210:** 227-237.

Oki M, Yamamoto H, Taniguchi H, Adachi Y, Imai K, Shinomura Y (2008). Overexpression of the receptor tyrosine kinase EphA4 in human gastric cancers. *World J Gastroenterol* **14:** 5650-5656.

Onichtchouk D, Chen YG, Dosch R, Gawantka V, Delius H, Massague J *et al.* (1999). Silencing of TGF-beta signalling by the pseudoreceptor BAMBI. *Nature* **401:** 480-485.

Orend G, Chiquet-Ehrismann R (2006). Tenascin-C induced signaling in cancer. *Cancer Lett* **244:** 143-163.

Oshima T, Akaike M, Yoshihara K, Shiozawa M, Yamamoto N, Sato T *et al.* (2008). Overexpression of EphA4 gene and reduced expression of EphB2 gene correlates with liver metastasis in colorectal cancer. *Int J Oncol* **33:** 573-577.

Parker SL, Tong T, Bolden S, Wingo PA (1997). Cancer statistics, 1997. *CA Cancer J Clin* **47:** 5-27.

Parkin DM, Bray F, Ferlay J, Pisani P (2005). Global cancer statistics, 2002. *CA Cancer J Clin* **55:** 74-108.

Ramaswamy S, Ross KN, Lander ES, Golub TR (2003). A molecular signature of metastasis in primary solid tumors. *Nat Genet* **33:** 49-54.

Remmele W (1996). *Pathologie 2, Verdauungstrakt*, 2 edn. Springer-Verlag, Berlin, Heidelberg, New York.

Rhodes DR, Yu J, Shanker K, Deshpande N, Varambally R, Ghosh D *et al.* (2004). ONCOMINE: a cancer microarray database and integrated data-mining platform. *Neoplasia* **6:** 1-6.

Ribeiro MM, Sarmento JA, Sobrinho Simoes MA, Bastos J (1981). Prognostic significance of Lauren and Ming classifications and other pathologic parameters in gastric carcinoma. *Cancer* **47:** 780-784.

Robert Koch Institut und Gesellschaft der epidemiologischen Krebsregister in Deutschland e.V. (2008). *Krebs in Deutschland 2003-2004, Häufigkeiten und Trends*, vol. 6.

Rorive S, Maris C, Debeir O, Sandras F, Vidaud M, Bieche I *et al.* (2006). Exploring the distinctive biological characteristics of pilocytic and low-grade diffuse astrocytomas using microarray gene expression profiles. *J Neuropathol Exp Neurol* **65:** 794-807.

Roukos D, Agnantis N, Fatouros M, Kappas A (2002). Gastric Cancer: Introduction, Pathology, Epidemiology. *Gastric Breast Cancer* **1:** 1-3.

Rubie C, Kempf K, Hans J, Su T, Tilton B, Georg T *et al.* (2005). Housekeeping gene variability in normal and cancerous colorectal, pancreatic, esophageal, gastric and hepatic tissues. *Mol Cell Probes* **19:** 101-109.

Ryu KW, Lee JH, Kim HS, Kim YW, Choi IJ, Bae JM (2003). Prediction of lymph nodes metastasis by sentinel node biopsy in gastric cancer. *Eur J Surg Oncol* **29:** 895-899.

Sadvakassova G, Dobocan MC, Difalco MR, Congote LF (2009). Regulator of differentiation 1 (ROD1) binds to the amphipathic C-terminal peptide of thrombospondin-4 and is involved in its mitogenic activity. *J Cell Physiol* **220:** 672-679.

Schwartz GK (1996). Invasion and metastases in gastric cancer: in vitro and in vivo models with clinical correlations. *Semin Oncol* **23:** 316-324.

Sekiya T, Oda T, Matsuura K, Akiyama T (2004). Transcriptional regulation of the TGF-beta pseudoreceptor BAMBI by TGF-beta signaling. *Biochem Biophys Res Commun* **320:** 680-684.

Shaw G, Morse S, Ararat M, Graham FL (2002). Preferential transformation of human neuronal cells by human adenoviruses and the origin of HEK 293 cells. *Faseb J* **16:** 869-871.

Shitashige M, Satow R, Honda K, Ono M, Hirohashi S, Yamada T (2008). Regulation of Wnt signaling by the nuclear pore complex. *Gastroenterology* **134:** 1961-1971, 1971 e1961-1964.

Siewert JR, Bottcher K, Stein HJ, Roder JD (1998). Relevant prognostic factors in gastric cancer: ten-year results of the German Gastric Cancer Study. *Ann Surg* **228:** 449-461.

Sobin LH, Wittekind C (eds) (1997) *UICC. TNM Classification of Malignant Tumours (5th edition)*. John Wiley & Sons: New York.

Sobin LH, Gospodarowicz MK, Wittekind C (eds) (2010) *UICC. TNM Classification of Malignant Tumours (7th edition)*. Wiley-Blackwell, 328pp.

Sorlie T, Tibshirani R, Parker J, Hastie T, Marron JS, Nobel A *et al.* (2003). Repeated observation of breast tumor subtypes in independent gene expression data sets. *Proc Natl Acad Sci U S A* **100:** 8418-8423.

Spicer J (2004). *Making Sense Of Multivariate Data Analysis*. SAGE Publications Inc.

Stemmermann GN, Brown C (1974). A survival study of intestinal and diffuse types of gastric carcinoma. *Cancer* **33:** 1190-1195.

Storey JD (2002). A direct approach to false discovery rates. *Journal Of The Royal Statistical Society Series B* **64:** 479-498.

Storey JD, Tibshirani R (2003). Statistical significance for genomewide studies. *Proc Natl Acad Sci U S A* **100:** 9440-9445.

Sugimura J, Yang XJ, Tretiakova MS, Takahashi M, Kort EJ, Fulton B *et al.* (2004). Gene expression profiling of mesoblastic nephroma and Wilms tumors--comparison and clinical implications. *Urology* **64:** 362-368; discussion 368.

Tahara E (2004). Genetic pathways of two types of gastric cancer. *IARC Sci Publ***:** 327-349.

Takeno A, Takemasa I, Seno S, Yamasaki M, Motoori M, Miyata H *et al.* (2010). Gene expression profile prospectively predicts peritoneal relapse after curative surgery of gastric cancer. *Ann Surg Oncol* **17:** 1033-1042.

Terashima M, Maesawa C, Oyama K, Ohtani S, Akiyama Y, Ogasawara S *et al.* (2005). Gene expression profiles in human gastric cancer: expression of maspin correlates with lymph node metastasis. *Br J Cancer* **92:** 1130-1136.

Terry LJ, Shows EB, Wente SR (2007). Crossing the nuclear envelope: hierarchical regulation of nucleocytoplasmic transport. *Science* **318:** 1412-1416.

Togo N, Ohwada S, Sakurai S, Toya H, Sakamoto I, Yamada T *et al.* (2008). Prognostic significance of BMP and activin membrane-bound inhibitor in colorectal cancer. *World J Gastroenterol* **14:** 4880-4888.

Tseng YH, Vicent D, Zhu J, Niu Y, Adeyinka A, Moyers JS *et al.* (2001). Regulation of growth and tumorigenicity of breast cancer cells by the low molecular weight GTPase Rad and nm23. *Cancer Res* **61:** 2071-2079.

Turashvili G, Bouchal J, Burkadze G, Kolar Z (2005). Differentiation of tumours of ductal and lobular origin: II. Genomics of invasive ductal and lobular breast carcinomas. *Biomed Pap Med Fac Univ Palacky Olomouc Czech Repub* **149:** 63-68.

Vauhkonen M, Vauhkonen H, Sipponen P (2006). Pathology and molecular biology of gastric cancer. *Best Pract Res Clin Gastroenterol* **20:** 651-674.

Wang CS, Lin KH, Chen SL, Chan YF, Hsueh S (2004). Overexpression of SPARC gene in human gastric carcinoma and its clinic-pathologic significance. *Br J Cancer* **91:** 1924-1930.

Watanabe H, Jass JR, Sobin LH (1990). *WHO. Histological Typing of Oesophageal and Gastric Tumours*, 2 edn. Springer-Verlag: Berlin.

Wayman J, Forman D, Griffin SM (2001). Monitoring the changing pattern of esophago-gastric cancer: data from a UK regional cancer registry. *Cancer Causes Control* **12:** 943-949.

Werner M, Becker KF, Keller G, Hofler H (2001). Gastric adenocarcinoma: pathomorphology and molecular pathology. *J Cancer Res Clin Oncol* **127:** 207-216.

Wu CM, Lee YS, Wang TH, Lee LY, Kong WH, Chen ES *et al.* (2006). Identification of differential gene expression between intestinal and diffuse gastric cancer using cDNA microarray. *Oncol Rep* **15:** 57-64.

Wu Z, Irizarry RA, Gentleman R, Murillo FM, Spencer F (2004). A Model Based Background Adjustment for Oligonucleotide Expression Arrays. *Johns Hopkins University, Dept of Biostatistics Working Papers* Working Paper 1.

Yashiro M, Chung YS, Nishimura S, Inoue T, Sowa M (1996). Fibrosis in the peritoneum induced by scirrhous gastric cancer cells may act as "soil" for peritoneal dissemination. *Cancer* **77:** 1668-1675.

Yasui W, Oue N, Aung PP, Matsumura S, Shutoh M, Nakayama H (2005). Molecular-pathological prognostic factors of gastric cancer: a review. *Gastric Cancer* **8:** 86-94.

Yokota T, Kunii Y, Teshima S, Yamada Y, Saito T, Takahashi M *et al.* (2000). Significant prognostic factors in patients with early gastric cancer. *Int Surg* **85:** 286-290.

Yokota T, Ishiyama S, Saito T, Teshima S, Narushima Y, Murata K *et al.* (2004). Lymph node metastasis as a significant prognostic factor in gastric cancer: a multiple logistic regression analysis. *Scand J Gastroenterol* **39:** 380-384.

Zhu J, Tseng YH, Kantor JD, Rhodes CJ, Zetter BR, Moyers JS *et al.* (1999). Interaction of the Ras-related protein associated with diabetes rad and the putative tumor metastasis suppressor NM23 provides a novel mechanism of GTPase regulation. *Proc Natl Acad Sci U S A* **96:** 14911-14918.

Zweig MH, Campbell G (1993). Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine. *Clin Chem* **39:** 561-577.

Appendix

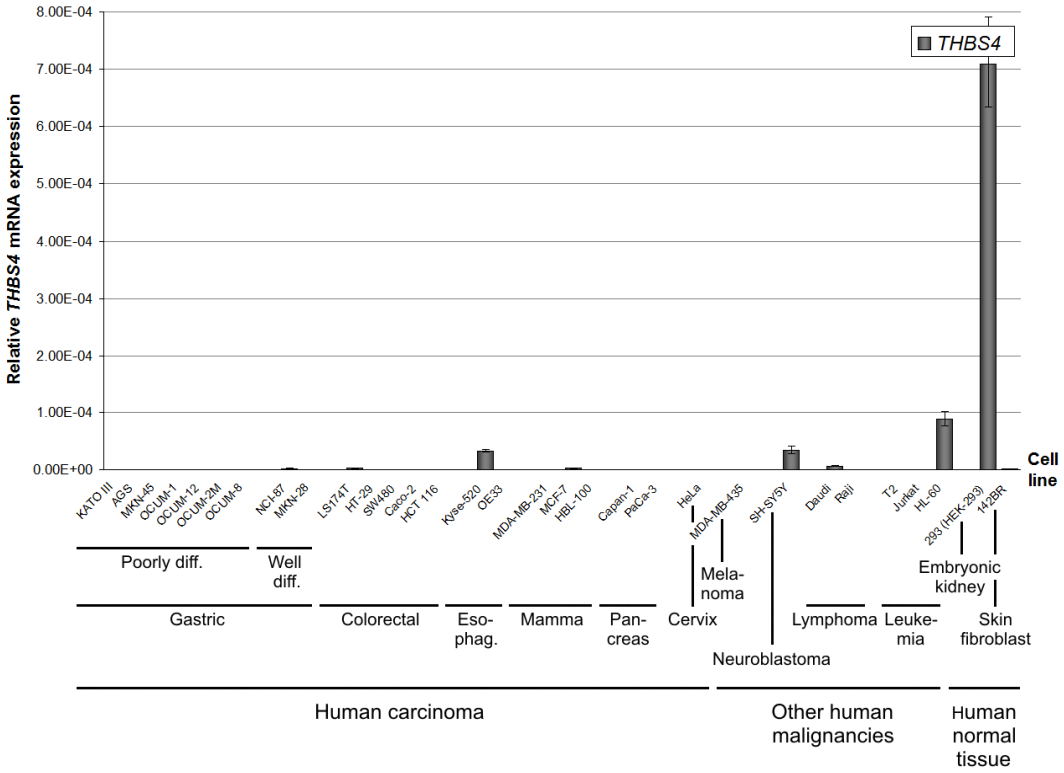


Figure 38: *THBS4* mRNA expression in different human cell lines.

The mRNA abundance was examined by means of quantitative real-time PCR in triplicates. Quantitation was done relative to the transcript of *GAPDH*. Error bars represent integrated standard errors of the mean.

diff. – differentiated; esophag. – esophageal

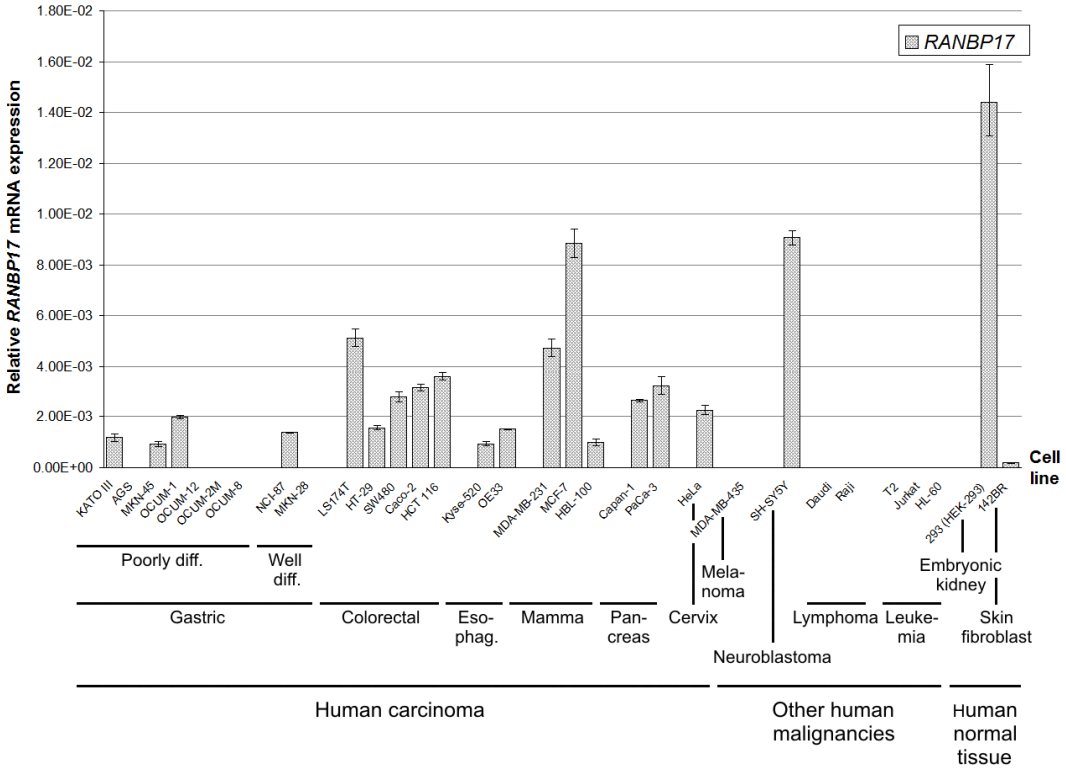


Figure 39: *RANBP17* mRNA expression in different human cell lines.

The mRNA abundance was examined by means of quantitative real-time PCR in triplicates. Quantitation was done relative to the transcript of *GAPDH*. Error bars represent integrated standard errors of the mean.

diff. – differentiated; esophag. – esophageal

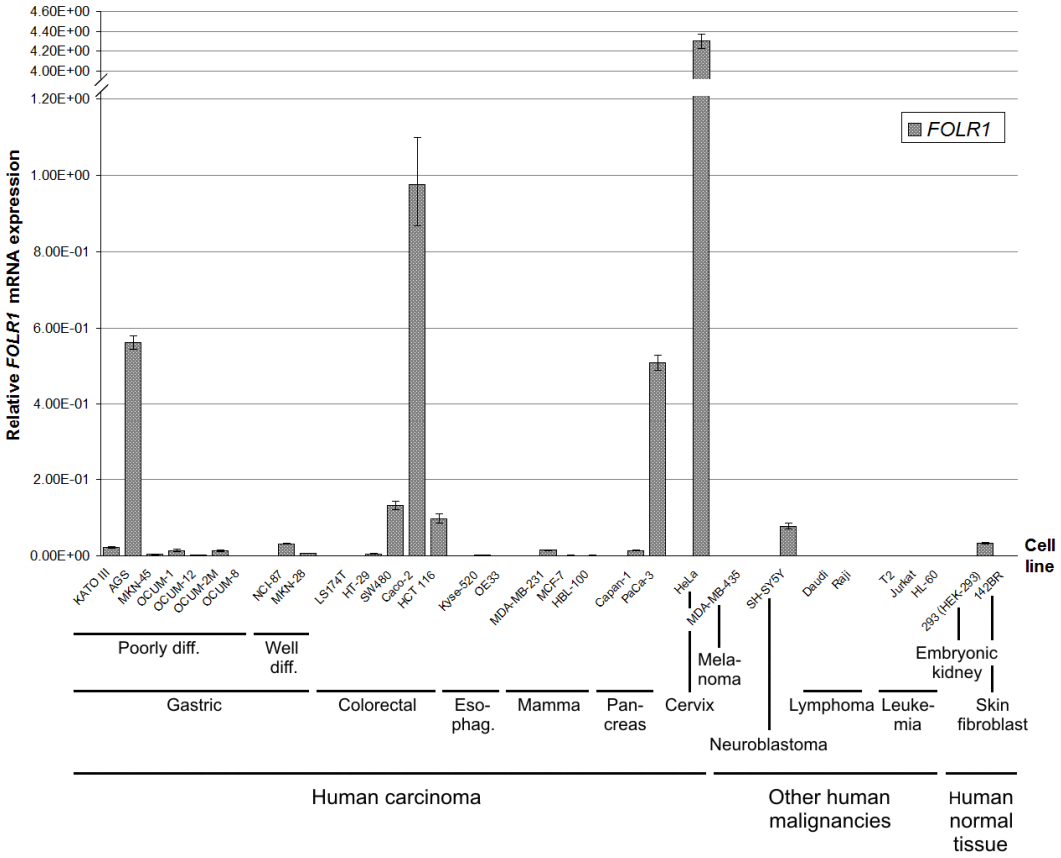


Figure 40: *FOLR1* mRNA expression in different human cell lines.

The mRNA abundance was examined by means of quantitative real-time PCR in triplicates. Quantitation was done relative to the transcript of *GAPDH*. Error bars represent integrated standard errors of the mean.

diff. – differentiated; esophag. – esophageal

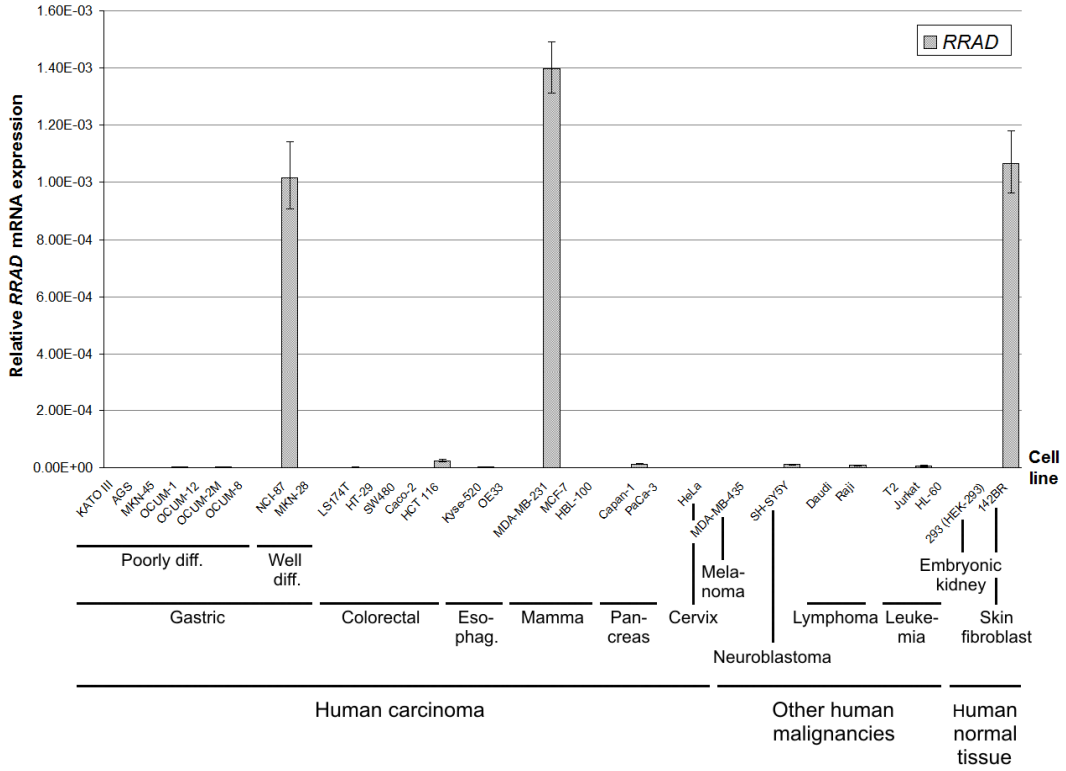


Figure 41: *RRAD* mRNA expression in different human cell lines.

The mRNA abundance was examined by means of quantitative real-time PCR in triplicates. Quantitation was done relative to the transcript of *GAPDH*. Error bars represent integrated standard errors of the mean.

diff. – differentiated; esophag. – esophageal

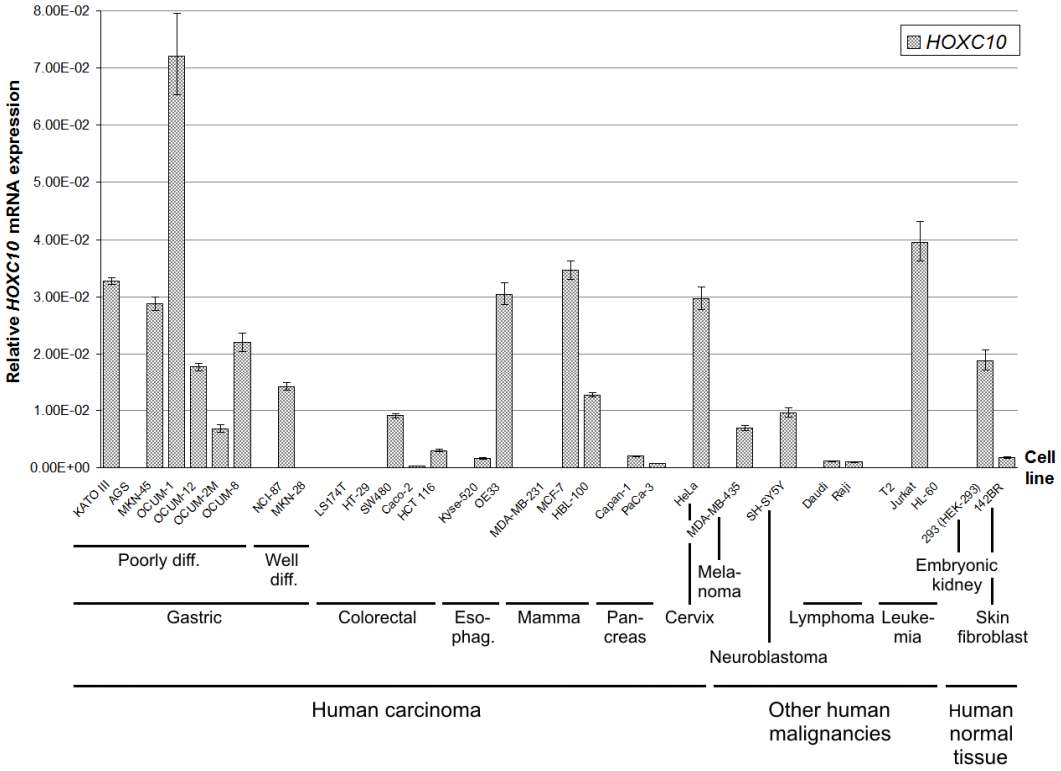


Figure 42: *HOXC10* mRNA expression in different human cell lines.

The mRNA abundance was examined by means of quantitative real-time PCR in triplicates. Quantitation was done relative to the transcript of *GAPDH*. Error bars represent integrated standard errors of the mean.

diff. – differentiated; esophag. – esophageal

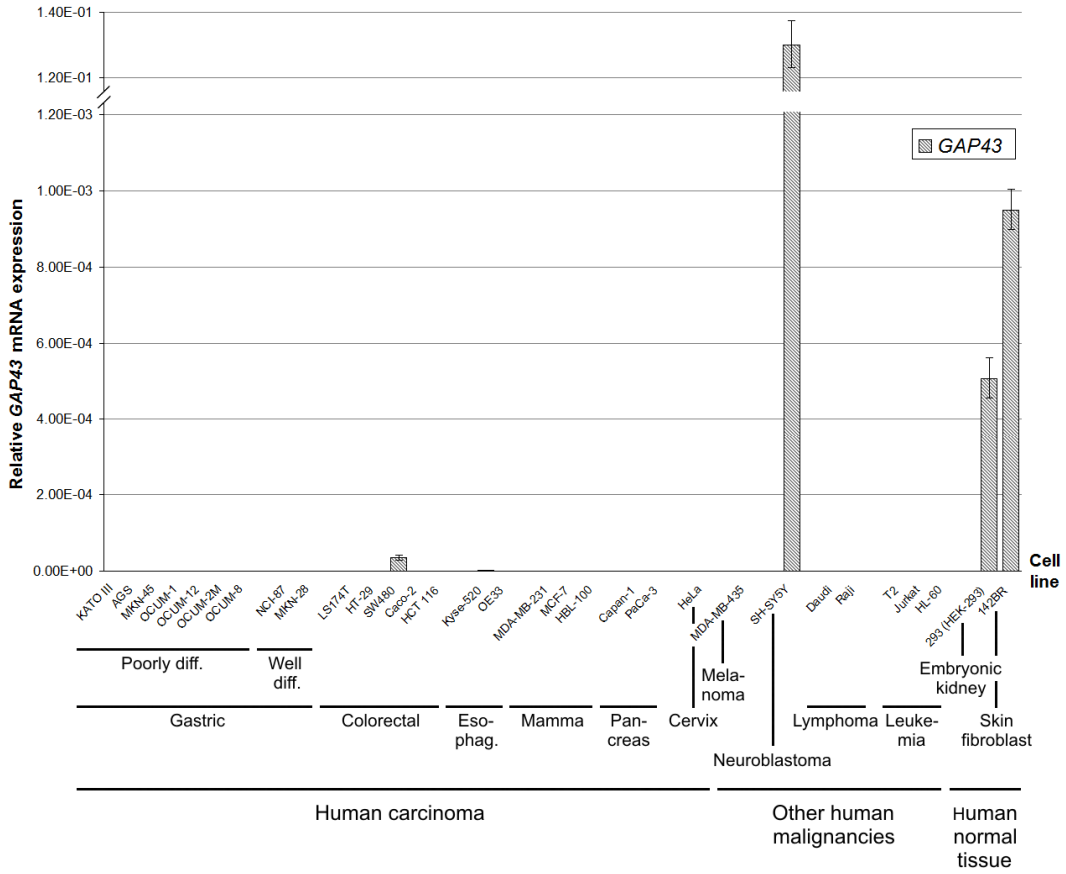


Figure 43: *GAP43* mRNA expression in different human cell lines.

The mRNA abundance was examined by means of quantitative real-time PCR in triplicates. Quantitation was done relative to the transcript of *GAPDH*. Error bars represent integrated standard errors of the mean.

diff. – differentiated; esophag. – esophageal

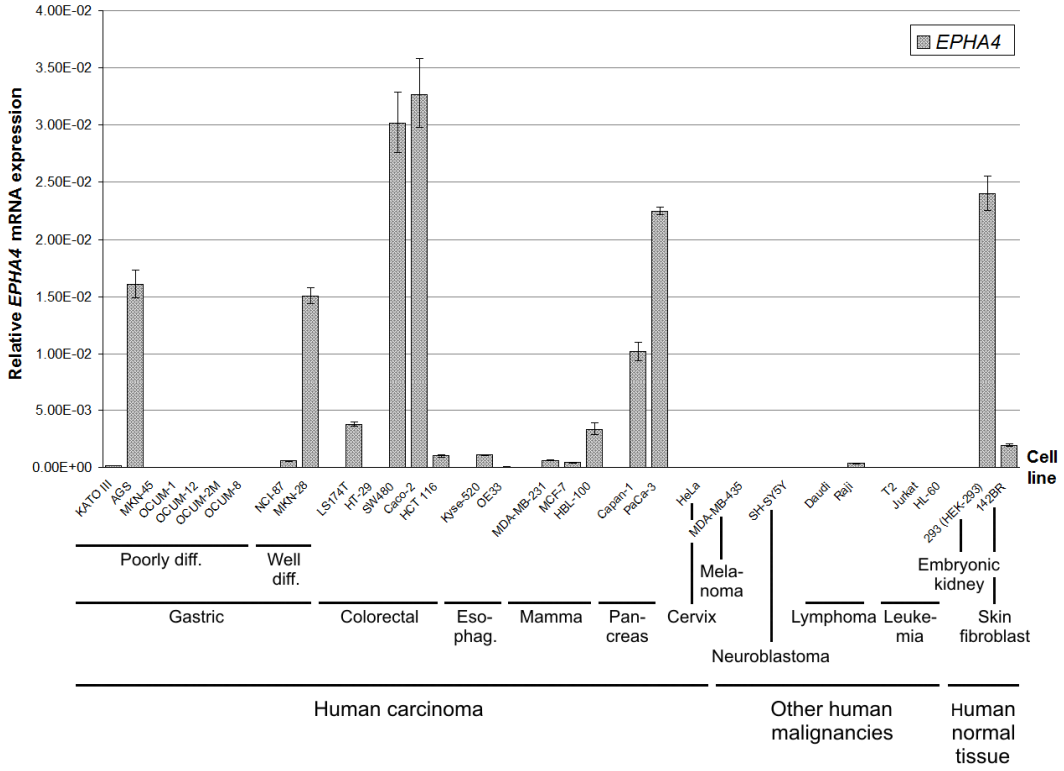


Figure 44: *EPHA4* mRNA expression in different human cell lines.

The mRNA abundance was examined by means of quantitative real-time PCR in triplicates. Quantitation was done relative to the transcript of *GAPDH*. Error bars represent integrated standard errors of the mean.

diff. – differentiated; esophag. – esophageal

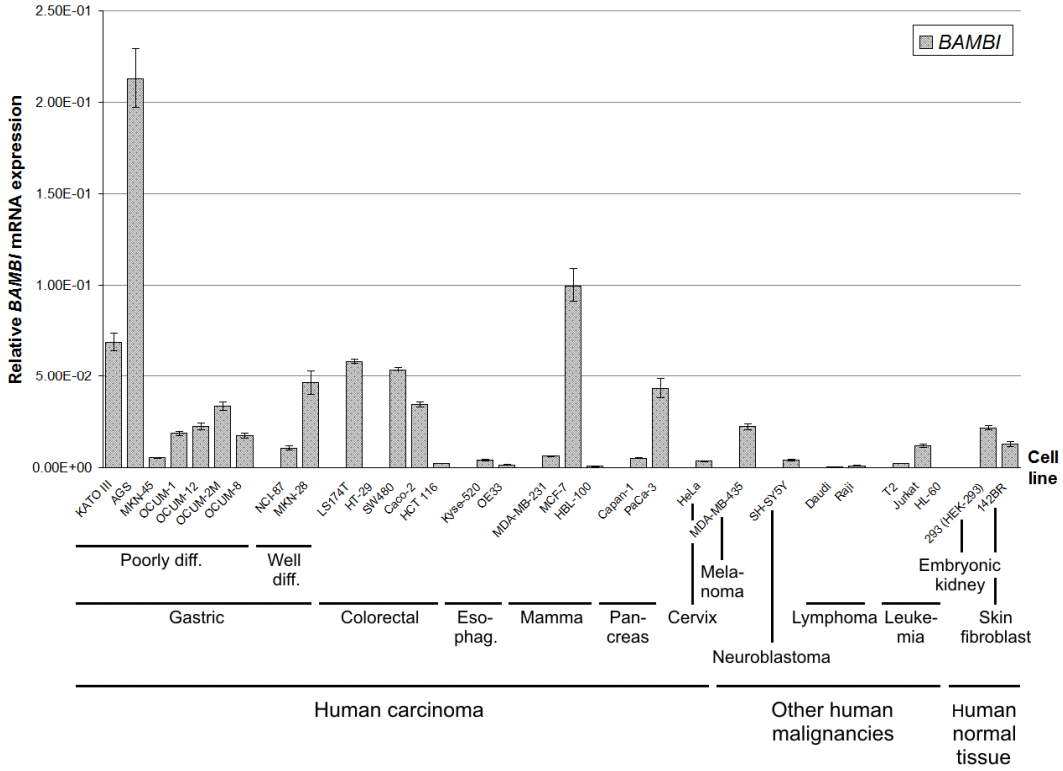


Figure 45: *BAMBI* mRNA expression in different human cell lines.

The mRNA abundance was examined by means of quantitative real-time PCR in triplicates. Quantitation was done relative to the transcript of *GAPDH*. Error bars represent integrated standard errors of the mean.

diff. – differentiated; esophag. – esophageal

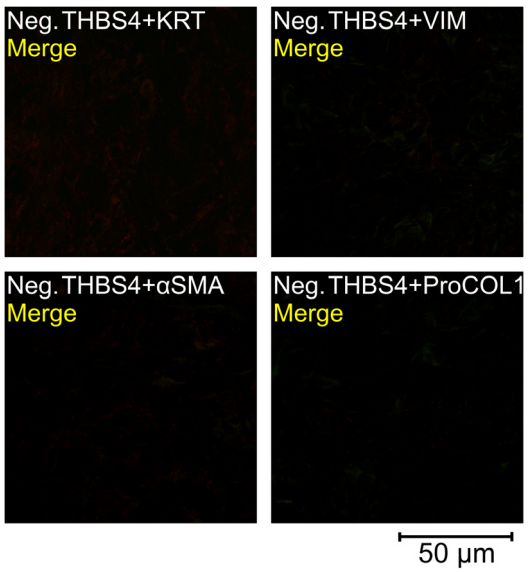


Figure 46: Negative controls for THBS4 colocalization studies.

Respective serial sections were incubated with fluorescently labeled secondary antibodies, only. Primary antibodies were omitted. Sections were scanned with identical settings (pinhole, excitation, frame average, *etc.*) to positive stainings, respectively.

Table 17: Annotation of the 50 genes with the most significant overexpression in human diffuse-type gastric adenocarcinomas compared to intestinal-type ones.

Genes significantly differentially expressed between diffuse and intestinal-type samples were extracted from microarray data using Welch-test with Benjamini and Hochberg FDR (corrected $p<0.05$; FC≥2). The 50 transcripts with the most significant overexpression in the diffuse type are depicted sorted with descending significance. If a transcript was represented by multiple probe sets, the one with the strongest significance is presented. Annotation information was obtained from NetAffx Analysis Center (09/2009).

corr. – corrected; FC – fold change; ID – identifier; GO – gene ontology; reg. – reg.; activ. – activation; prolif. – proliferation; bind. – binding; pos. – positive; neg. – negative

Gene symbol	Gene title	Corr. *p*-value	FC	Probe set ID	UniGene ID	GO biological process term	GO molecular function term	GO cellular component term
THBS4	thrombospondin 4	1.65E-07	40.8	204776_at	Hs.211426	substrate-bound cell migration, cell extension /// cell adhesion	structural molecule activity /// calcium ion bind. /// protein bind. /// heparin bind.	extracellular region /// proteinaceous extracellular matrix /// platelet alpha granule lumen
RHOJ	ras homolog gene family, member J	1.65E-07	6.6	238906_s_at	Hs.656339	small GTPase mediated signal transduction /// Rho protein signal transduction /// reg. of cell shape /// actin cytoskeleton organization	nucleotide bind. /// GTPase activity /// protein bind. /// GTP bind.	intracellular /// plasma membrane /// membrane
TSHZ2	teashirt zinc finger homeobox 2	2.95E-07	5.8	238577_s_at	Hs.649877	transcription /// reg. of transcription, DNA-dependent /// multicellular organismal development	DNA bind. /// transcription factor activity /// zinc ion bind. /// sequence-specific DNA bind. /// metal ion bind.	extracellular region /// proteinaceous extracellular matrix /// intracellular /// nucleus /// extracellular matrix
BHMT2	betaine-homocysteine methyltransferase 2	1.04E-06	4.6	219902_at	Hs.114172	methionine biosynthetic process	methyltransferase activity /// zinc ion bind. /// homocysteine S-methyltransferase activity /// transferase activity /// metal ion bind. /// betaine-homocysteine S-methyltransferase activity	cytoplasm
SPARCL1	SPARC-like 1 (hevin)	1.63E-06	3.5	200795_at	Hs.62886	signal transduction	calcium ion bind.	extracellular region /// proteinaceous extracellular matrix /// synapse
EDNRA	endothelin receptor type A	1.63E-06	3.5	204464_s_at	Hs.183713	smooth muscle contraction /// signal transduction /// G-protein coupled receptor protein signaling pathway /// activ. of adenylate cyclase activity /// activ. of phospholipase C activity /// elevation of cytosolic calcium ion concentration /// respiratory gaseous exchange /// cell prolif. /// artery smooth muscle contraction /// glucose transport /// vasoconstriction	endothelin-A receptor activity /// phosphoinositide phospholipase C activity /// signal transducer activity /// receptor activity /// G-protein coupled receptor activity /// endothelin receptor activity /// endothelin receptor activity /// protein bind.	plasma membrane /// integral to plasma membrane /// membrane /// integral to membrane
C1R	complement component 1, r subcomponent	1.63E-06	2.9	212067_s_at	Hs.524224	proteolysis /// immune response /// complement activ., classical pathway /// innate immune response	catalytic activity /// serine-type endopeptidase activity /// calcium ion bind. /// peptidase activity /// serine-type peptidase activity /// hydrolase activity	extracellular region
---	---	1.63E-06	4.1	230309_at	Hs.651174	---	---	---

Gene symbol	Gene title	Corr. *p*-value	FC	Probe set ID	UniGene ID	GO biological process term	GO molecular function term	GO cellular component term
PBX3	pre-B-cell leukemia homeobox 3	2.20E-06	3.8	204082_at	Hs.428027	reg. of respiratory gaseous exchange by neurological process /// transcription /// reg. of transcription, DNA-dependent /// anterior compartment specification /// posterior compartment specification /// respiratory gaseous exchange /// adult locomotory behavior /// dorsal spinal cord development /// reg. of transcription /// neuron development	DNA bind. /// transcription factor activity /// protein bind. /// transcription regulator activity /// sequence-specific DNA bind.	nucleus /// transcription factor complex /// cytoplasm
LAMA2	laminin, alpha 2	2.43E-06	4.1	213519_s_at	Hs.200841	cell adhesion /// muscle organ development /// reg. of cell adhesion /// reg. of cell migration /// pos. reg. of synaptic transmission, cholinergic /// reg. of embryonic development	receptor bind. /// structural molecule activity /// protein bind.	extracellular region /// proteinaceous extracellular matrix /// basement membrane /// basal lamina /// laminin-1 complex /// extracellular matrix /// sarcolemma
CCDC80	coiled-coil domain containing 80	3.32E-06	12.4	243864_at	Hs.477128	---	---	extracellular region /// proteinaceous extracellular matrix /// viral envelope /// intracellular membrane-bounded organelle
ERG	v-ets erythroblastosis virus E26 oncogene homolog (avian)	3.32E-06	3.7	213541_s_at	Hs.473819	transcription /// reg. of transcription, DNA-dependent /// protein amino acid phosphorylation /// signal transduction /// multicellular organismal development /// cell prolif.	DNA bind. /// transcription factor activity /// signal transducer activity /// protein bind. /// sequence-specific DNA bind.	nucleus
HEYL	hairy/enhancer-of-split related with YRPW motif-like	3.77E-06	4.9	226828_s_at	Hs.472566	transcription /// reg. of transcription, DNA-dependent /// Notch signaling pathway /// multicellular organismal development /// nervous system development /// reg. of transcription	DNA bind. /// transcription factor activity /// protein bind. /// transcription regulator activity	nucleus
---	---	3.77E-06	4.8	227618_at	Hs.496303	---	---	---
SHISA3	shisa homolog 3 (Xenopus laevis)	3.77E-06	14.4	229485_x_at	Hs.370904	multicellular organismal development	---	endoplasmic reticulum /// endoplasmic reticulum membrane /// membrane /// integral to membrane
JAM2	junctional adhesion molecule 2	4.45E-06	4.5	219213_at	Hs.517227	cell-cell adhesion	---	plasma membrane /// integral to plasma membrane /// tight junction /// membrane /// integral to membrane /// cell junction
GUCY1B3	guanylate cyclase 1, soluble, beta 3	4.57E-06	3.2	203817_at	Hs.77890	cGMP biosynthetic process /// intracellular signaling cascade /// nitric oxide mediated signal transduction /// blood circulation /// cyclic nucleotide biosynthetic process	guanylate cyclase activity /// guanylate cyclase activity /// receptor activity /// iron ion bind. /// lyase activity /// phosphorus-oxygen lyase activity /// heme bind. /// metal ion bind.	cytoplasm /// guanylate cyclase complex, soluble
ZCCHC24	zinc finger, CCHC domain containing 24	4.91E-06	3.7	212423_at	Hs.523080	---	nucleic acid bind. /// zinc ion bind. /// metal ion bind.	---

Gene symbol	Gene title	Corr. *p*-value	FC	Probe set ID	UniGene ID	GO biological process term	GO molecular function term	GO cellular component term
IGFBP7	insulin-like growth factor bind. protein 7	5.10E-06	3.1	213910_at	Hs.479808	reg. of cell growth /// cell adhesion /// neg. reg. of cell prolif.	protein bind. /// insulin-like growth factor bind. /// growth factor bind.	extracellular region
BOC	Boc homolog (mouse)	5.14E-06	5.6	225990_at	Hs.591318	cell adhesion /// pos. reg. of myoblast differentiation	protein bind.	plasma membrane /// membrane /// integral to membrane
GGT5	gamma-glutamyltransferase 5	5.35E-06	3.6	205582_s_at	Hs.437156	amino acid metabolic process /// glutathione metabolic process /// glutathione biosynthetic process /// leukotriene biosynthetic process	gamma-glutamyltransferase activity /// gamma-glutamyltransferase activity /// acyltransferase activity /// transferase activity	membrane /// integral to membrane
DDR2	discoidin domain receptor tyrosine kinase 2	5.35E-06	3.9	225442_at	Hs.275757	protein amino acid phosphorylation /// cell adhesion /// signal transduction /// transmembrane receptor protein tyrosine kinase signaling pathway /// pos. reg. of cell prolif.	nucleotide bind. /// transmembrane receptor protein tyrosine kinase activity /// receptor activity /// ATP bind. /// transferase activity	integral to plasma membrane /// membrane
---	---	5.35E-06	4.1	232935_at	Hs.659578	---	---	---
BNC2	basonuclin 2	5.82E-06	8.4	220272_at	Hs.656581	transcription /// reg. of transcription, DNA-dependent	zinc ion bind. /// metal ion bind.	intracellular /// nucleus
C10orf72	chromosome 10 open reading frame 72	7.59E-06	4.1	235471_at	Hs.522928	---	---	plasma membrane /// membrane /// integral to membrane
PBX1	pre-B-cell leukemia homeobox 1	7.74E-06	3.2	212148_at	Hs.557097	transcription /// reg. of transcription, DNA-dependent /// C21-steroid hormone biosynthetic process /// sex determination /// cell differentiation /// reg. of transcription	DNA bind. /// transcription factor activity /// protein bind. /// transcription regulator activity /// sequence-specific DNA bind.	nucleus /// nucleolus /// cytoplasm
PDE1A	phosphodiesterase 1A, calmodulin-dependent	8.34E-06	5.2	236234_at	---	signal transduction	calmodulin-dependent cyclic-nucleotide phosphodiesterase activity /// calmodulin bind. /// hydrolase activity	---
HGF	hepatocyte growth factor (hepapoietin A; scatter factor)	8.92E-06	3.7	209960_at	Hs.396530	activ. of MAPK activity /// cell morphogenesis /// epithelial to mesenchymal transition /// liver development /// proteolysis /// anti-apoptosis /// mitosis /// cell prolif. /// hepatocyte growth factor receptor signaling pathway /// myoblast prolif.	catalytic activity /// serine-type endopeptidase activity /// protein bind. /// growth factor activity	extracellular region /// platelet alpha granule lumen
CRISPLD1	cysteine-rich secretory protein LCCL domain containing 1	8.92E-06	9.8	223475_at	Hs.436542	---	---	extracellular region
ANKRD35	ankyrin repeat domain 35	8.92E-06	3.2	231118_at	Hs.710624	---	---	---
SSBP2	single-stranded DNA bind. protein 2	8.97E-06	2.6	203787_at	Hs.102735	reg. of transcription	DNA bind. /// single-stranded DNA bind. /// protein bind. /// transcription regulator activity	nucleus /// cytoplasm

Gene symbol	Gene title	Corr. *p*-value	FC	Probe set ID	UniGene ID	GO biological process term	GO molecular function term	GO cellular component term
OMD	osteomodulin	8.99E-06	14.7	205908_s_at	Hs.94070	cell adhesion	protein bind.	extracellular region /// proteinaceous extracellular matrix
PRICKLE1	prickle homolog 1 (Drosophila)	9.09E-06	4.7	230708_at	Hs.524348	---	zinc ion bind. /// metal ion bind.	nucleus /// cytoplasm /// cytosol /// membrane /// nuclear membrane
DCLK1	doublecortin-like kinase 1	9.13E-06	15.2	229800_at	Hs.507755	protein amino acid phosphorylation /// intracellular signaling cascade /// multicellular organismal development /// nervous system development /// central nervous system development /// endosome transport /// cell differentiation	nucleotide bind. /// protein kinase activity /// protein serine/threonine kinase activity /// receptor signaling protein activity /// protein bind. /// ATP bind. /// kinase activity /// transferase activity	integral to plasma membrane
MMRN2	multimerin 2	9.65E-06	3.0	236262_at	Hs.719260	---	---	extracellular region /// proteinaceous extracellular matrix /// extracellular space
LHFP	lipoma HMGIC fusion partner	1.03E-05	3.3	218656_s_at	Hs.507798	---	DNA bind.	membrane /// integral to membrane
C21orf34	chromosome 21 open reading frame 34	1.17E-05	8.3	239999_at	Hs.719301	---	---	---
GNG11	guanine nucleotide bind. protein (G protein), gamma 11	1.24E-05	2.8	204115_at	Hs.83381	signal transduction /// G-protein coupled receptor protein signaling pathway /// hormone-mediated signaling	GTPase activity /// signal transducer activity	heterotrimeric G-protein complex /// plasma membrane /// membrane
SSPN	sarcospan (Kras oncogene-associated gene)	1.26E-05	3.9	226932_at	Hs.183428	muscle contraction /// cell adhesion	---	plasma membrane /// integral to plasma membrane /// dystrophin-associated glycoprotein complex /// membrane /// integral to membrane /// cell junction /// transport vesicle /// synapse /// postsynaptic membrane
FBXL7	F-box and leucine-rich repeat protein 7	1.30E-05	2.8	213249_at	Hs.433057	ubiquitin-dependent protein catabolic process /// modification-dependent protein catabolic process	ubiquitin-protein ligase activity /// protein bind.	ubiquitin ligase complex
MGP	matrix Gla protein	1.40E-05	6.7	202291_s_at	Hs.365706	cartilage condensation /// ossification /// protein complex assembly /// multicellular organismal development /// response to nutrient /// response to mechanical stimulus /// response to hormone stimulus /// cell differentiation /// lung development /// reg. of bone mineralization /// branching morphogenesis of a tube /// cartilage development /// response to calcium ion	extracellular matrix structural constituent /// calcium ion bind. /// structural constituent of bone /// calcium-dependent protein bind.	extracellular region /// proteinaceous extracellular matrix /// extracellular space /// endoplasmic reticulum /// extracellular matrix /// protein complex
---	---	1.40E-05	5.7	229127_at	Hs.655065	---	---	---

Gene symbol	Gene title	Corr. *p*-value	FC	Probe set ID	UniGene ID	GO biological process term	GO molecular function term	GO cellular component term
PDGFRB	platelet-derived growth factor receptor, beta polypeptide	1.41E-05	2.7	202273_at	Hs.509067	in utero embryonic development /// kidney development /// tissue homeostasis /// protein amino acid phosphorylation /// nitrogen compound metabolic process /// signal transduction /// transmembrane receptor protein tyrosine kinase signaling /// pathway /// pos. reg. of cell prolif. /// peptidyl-tyrosine phosphorylation /// protein amino acid autophosphorylation /// platelet-derived growth factor receptor signaling pathway /// vascular endothelial growth factor receptor signaling pathway /// skeletal system morphogenesis /// smooth muscle tissue development /// reg. of peptidyl-tyrosine phosphorylation /// cell chemotaxis ...	nucleotide bind. /// protein kinase activity /// protein tyrosine kinase activity /// transmembrane receptor protein tyrosine kinase activity /// signal transducer activity /// receptor activity /// platelet activating factor receptor activity /// platelet-derived growth factor receptor activity /// platelet-derived growth factor beta-receptor activity /// vascular endothelial growth factor receptor activity /// platelet-derived growth factor receptor bind. /// protein bind. /// ATP bind. /// kinase activity /// transferase activity /// growth factor bind. /// platelet-derived growth factor bind.	integral to plasma membrane /// membrane /// integral to membrane
MPDZ	multiple PDZ domain protein	1.41E-05	3.9	205079_s_at	Hs.169378	interspecies interaction between organisms	protein bind.	plasma membrane /// tight junction /// postsynaptic density /// membrane /// synaptosome /// cell junction /// dendrite /// cell projection /// synapse /// postsynaptic membrane
GHR	growth hormone receptor	1.47E-05	8.0	205498_at	Hs.125180	activ. of MAPK activity /// allantoin metabolic process /// citrate metabolic process /// 2-oxoglutarate metabolic process /// succinate metabolic process /// oxaloacetate metabolic process /// isoleucine metabolic process /// JAK-STAT cascade /// hormone-mediated signaling // response to morphine /// cartilage development involved in endochondral bone morphogenesis /// response to interleukin-1 ...	receptor activity /// cytokine receptor activity /// growth hormone receptor activity /// protein bind. /// peptide hormone bind. /// protein kinase bind. /// protein phosphatase bind. /// SH2 domain bind. /// phosphate bind. /// protein homodimerization activity /// proline-rich region bind. ...	extracellular region /// extracellular space /// nucleus /// cytoplasm /// mitochondrion /// plasma membrane /// integral to plasma membrane /// cell surface /// membrane /// integral to membrane /// extrinsic to membrane /// cell soma /// receptor complex /// growth hormone receptor complex
LAMA4	laminin, alpha 4	1.50E-05	2.5	202202_s_at	Hs.654572	blood vessel development /// cell adhesion /// reg. of cell adhesion /// reg. of cell migration /// reg. of embryonic development /// brown fat cell differentiation	receptor bind. /// extracellular matrix structural constituent /// protein bind.	extracellular region /// proteinaceous extracellular matrix /// basement membrane /// basal lamina /// laminin-1 complex
DARC	Duffy blood group, chemokine receptor	1.53E-05	9.0	208335_s_at	Hs.153381	defense response /// inflammatory response /// signal transduction /// G-protein coupled receptor protein signaling pathway	signal transducer activity /// receptor activity /// transmembrane receptor activity /// G-protein coupled receptor activity /// chemokine receptor activity	plasma membrane /// membrane /// integral to membrane

Gene symbol	Gene title	Corr. *p*-value	FC	Probe set ID	UniGene ID	GO biological process term	GO molecular function term	GO cellular component term
DZIP1	DAZ interacting protein 1	1.77E-05	4.2	204557_s_at	Hs.656580	multicellular organismal development /// germ cell development /// spermatogenesis /// cell differentiation	protein bind. /// zinc ion bind. /// metal ion bind.	intracellular /// nucleus /// cytoplasm /// protein complex
LOC399959	hypothetical LOC399959	1.87E-05	6.3	225381_at	Hs.44098	---	---	---
---	---	1.87E-05	4.6	241457_at	Hs.661374	---	---	---

Table 18: Annotation of the 50 genes with the most significant overexpression in human intestinal-type gastric adenocarcinomas compared to diffuse-type ones.

Genes significantly differentially expressed between diffuse and intestinal-type samples were extracted from microarray data using Welch-test with Benjamini and Hochberg FDR (corrected $p<0.05$; FC≥2). The 50 transcripts with the most significant overexpression in the intestinal type are depicted sorted with descending significance. If a transcript was represented by multiple probe sets, the one with the strongest significance is presented. Annotation information was obtained from NetAffx Analysis Center (09/2009).

corr. – corrected; FC – fold change; ID – identifier; GO – gene ontology; reg. – reg.; activ. – activation; prolif. – proliferation; bind. – binding; pos. – positive; neg. – negative

Gene symbol	Gene title	Corr. *p*-value	FC	Probe set ID	UniGene ID	GO biological process term	GO molecular function term	GO cellular component term
SQLE	squalene epoxidase	8.97E-06	4.2	213577_at	Hs.71465	metabolic process /// sterol biosynthetic process /// oxidation reduction	monooxygenase activity /// squalene monooxygenase activity /// oxidoreductase activity /// FAD bind.	endoplasmic reticulum /// endoplasmic reticulum membrane /// microsome /// membrane /// integral to membrane
IDI1	isopentenyl-diphosphate delta isomerase 1	2.19E-05	2.2	208881_x_at	Hs.283652	steroid biosynthetic process /// cholesterol biosynthetic process /// isoprenoid biosynthetic process /// lipid biosynthetic process /// carotenoid biosynthetic process /// sterol biosynthetic process	magnesium ion bind. /// isopentenyl-diphosphate delta-isomerase activity /// hydrolase activity /// isomerase activity /// metal ion bind.	peroxisome /// cytosol
TMEM177	transmembrane protein 177	3.69E-05	2.3	218897_at	Hs.439991	---	---	membrane /// integral to membrane
HIST1H2BD	histone cluster 1, H2bd	3.84E-05	3.0	209911_x_at	Hs.591797	nucleosome assembly /// defense response to bacterium	DNA bind. /// protein bind.	nucleosome /// nucleus /// chromosome
CKS1B	CDC28 protein kinase regulatory subunit 1B	4.05E-05	2.2	201897_s_at	Hs.374378	reg. of cyclin-dependent protein kinase activity /// cell cycle /// cell prolif. /// cell division	protein bind. /// kinase activity /// cyclin-dependent protein kinase regulator activity	nucleoplasm
TRIB3	tribbles homolog 3 (Drosophila)	5.37E-05	6.4	218145_at	Hs.516826	transcription /// reg. of transcription, DNA-dependent /// protein amino acid phosphorylation /// neg. reg. of protein kinase activity /// apoptosis /// response to stress /// reg. of MAP kinase activity	transcription corepressor activity /// protein kinase activity /// protein kinase inhibitor activity /// protein bind. /// ATP bind. /// kinase activity /// protein kinase bind.	nucleus
HIST1H2AG	histone cluster 1, H2ag	5.77E-05	2.4	207156_at	Hs.51011	nucleosome assembly	DNA bind.	nucleosome /// nucleus /// chromosome

Gene symbol	Gene title	Corr. *p*-value	FC	Probe set ID	UniGene ID	GO biological process term	GO molecular function term	GO cellular component term
ACAT2	acetyl-Coenzyme A acetyltransferase 2	5.85E-05	2.5	209608_s_at	Hs.571037	lipid metabolic process /// metabolic process	catalytic activity /// acetyl-CoA C-acetyltransferase activity /// protein bind. /// acyltransferase activity /// transferase activity	cytoplasm
MRE11A	MRE11 meiotic recombination 11 homolog A (S. cerevisiae)	6.07E-05	2.1	205395_s_at	Hs.192649	reg. of mitotic recombination /// DNA metabolic process /// DNA repair /// double-strand break repair /// double-strand break repair via nonhomologous end joining /// DNA recombination /// response to DNA damage stimulus /// telomere maintenance via telomerase meiosis /// reciprocal meiotic recombination	single-stranded DNA specific endodeoxyribonuclease activity /// DNA bind. /// double-stranded DNA bind. /// nuclease activity /// endonuclease activity /// endodeoxyribonuclease activity /// exonuclease activity /// protein bind. /// protein C-terminus bind. /// 3'-5' exonuclease activity /// hydrolase activity /// manganese ion bind.	nucleus /// nucleoplasm /// nucleolus
---	---	6.52E-05	4.4	206110_at	---	nucleosome assembly	DNA bind. /// protein bind.	nucleosome /// nucleus /// chromosome
DARS2	aspartyl-tRNA synthetase 2, mitochondrial	6.78E-05	2.1	218365_s_at	Hs.647707	translation /// tRNA aminoacylation for protein translation /// aspartyl-tRNA aminoacylation	nucleotide bind. /// nucleic acid bind. /// aminoacyl-tRNA ligase activity /// aspartate-tRNA ligase activity /// ATP bind. /// ligase activity	cytoplasm /// mitochondrion /// mitochondrial matrix
DHCR7	7-dehydrocholesterol reductase	8.91E-05	2.7	201791_s_at	Hs.503134	blood vessel development /// steroid biosynthetic process /// cholesterol biosynthetic process /// lipid biosynthetic process /// post-embryonic development /// sterol biosynthetic process /// cell differentiation /// lung development /// multicellular organism growth /// reg. of cell prolif. /// oxidation reduction	protein bind. /// oxidoreductase activity /// 7-dehydrocholesterol reductase activity /// 7-dehydrocholesterol reductase activity /// 7-dehydrocholesterol reductase activity	nuclear outer membrane /// endoplasmic reticulum /// endoplasmic reticulum membrane /// membrane /// integral to membrane
PABPC1L	poly(A) bind. protein, cytoplasmic 1-like	9.68E-05	3.3	226670_s_at	Hs.641481	---	nucleotide bind. /// nucleic acid bind. /// RNA bind.	---
TOP2A	topoisomerase (DNA) II alpha 170kDa	9.95E-05	5.7	201291_s_at	Hs.156346	DNA metabolic process /// DNA replication /// DNA topological change /// DNA ligation /// DNA repair /// response to DNA damage stimulus /// chromosome segregation /// apoptotic chromosome condensation /// pos. reg. of apoptosis /// pos. reg. of apoptosis /// pos. reg. of viral genome replication /// pos. reg. of retroviral genome replication /// phosphoinositide-mediated signaling	nucleotide bind. /// DNA bind. /// chromatin bind. /// DNA topoisomerase activity /// DNA topoisomerase (ATP-hydrolyzing) activity /// protein kinase C bind. /// protein bind. /// ATP bind. /// protein C-terminus bind. /// DNA-dependent ATPase activity /// drug bind. /// isomerase activity /// enzyme bind. /// protein homodimerization activity /// histone deacetylase bind. /// ubiquitin bind. /// protein heterodimerization activity	nucleus /// nucleoplasm /// chromosome /// nucleolus /// cytoplasm /// centriole /// DNA topoisomerase complex (ATP-hydrolyzing) /// viral integration complex /// protein complex

Gene symbol	Gene title	Corr. *p*-value	FC	Probe set ID	UniGene ID	GO biological process term	GO molecular function term	GO cellular component term
RMND5A	required for meiotic nuclear division 5 homolog A (S. cerevisiae)	1.16E-04	2.2	212479_s_at	Hs.75277	---	---	---
PPIL1	peptidylprolyl isomerase (cyclophilin)-like 1	1.25E-04	2.1	222500_at	Hs.27693	mRNA processing /// protein folding /// RNA splicing	peptidyl-prolyl cis-trans isomerase activity /// isomerase activity	spliceosome
HIST1H2BH	histone cluster 1, H2bh	1.27E-04	3.4	208546_x_at	Hs.247815	nucleosome assembly	DNA bind.	nucleosome /// nucleus /// chromosome
---	---	1.38E-04	2.3	241838_at	Hs.664732	---	---	---
FAM54A	family with sequence similarity 54, member A	1.40E-04	2.4	234944_s_at	Hs.121536	---	---	---
SNORA21	small nucleolar RNA, H/ACA box 21	1.61E-04	3.5	215224_at	Hs.719782	---	---	---
CCDC138	coiled-coil domain containing 138	1.65E-04	2.3	235644_at	Hs.362702	---	---	---
XPO5	exportin 5	1.65E-04	2.0	223057_s_at	Hs.203206	protein export from nucleus /// transport /// protein transport /// gene silencing by RNA	tRNA bind. /// RNA bind. /// bind. /// protein bind. /// protein transporter activity	nucleus /// cytoplasm
DDX52	DEAD (Asp-Glu-Ala-Asp) box polypeptide 52	1.81E-04	2.4	210320_s_at	Hs.590937	---	nucleotide bind. /// nucleic acid bind. /// RNA bind. /// helicase activity /// ATP bind. /// ATP-dependent helicase activity /// hydrolase activity	nucleus
HIST1H2BE	histone cluster 1, H2be	1.83E-04	2.3	208527_x_at	Hs.534369	nucleosome assembly /// defense response to bacterium	DNA bind. /// protein bind.	nucleosome /// nucleus /// chromosome
C10orf2	chromosome 10 open reading frame 2	1.93E-04	2.7	218590_at	Hs.22678	DNA replication /// cell death	nucleotide bind. /// protease bind. /// DNA helicase activity /// helicase activity /// ATP bind. /// hydrolase activity	mitochondrion /// mitochondrial nucleoid
SUV39H2	suppressor of variegation 3-9 homolog 2 (Drosophila)	1.96E-04	2.4	1554572_a_at	Hs.554883	chromatin assembly or disassembly /// chromatin remodeling /// transcription /// reg. of transcription, DNA-dependent /// cell cycle /// chromatin modification /// cell differentiation	chromatin bind. /// protein bind. /// methyltransferase activity /// zinc ion bind. /// transferase activity /// histone-lysine N-methyltransferase activity /// histone methyltransferase activity (H3-K9 specific) /// histone methyltransferase activity (H3-K9 specific)	chromosome, centromeric region /// chromatin // nucleus /// chromosome
C1orf135	chromosome 1 open reading frame 135	2.01E-04	4.0	220011_at	Hs.149305	---	---	---

Gene symbol	Gene title	Corr. *p*-value	FC	Probe set ID	UniGene ID	GO biological process term	GO molecular function term	GO cellular component term
KIFC1	kinesin family member C1	2.50E-04	2.6	209680_s_at	Hs.436912	mitotic sister chromatid segregation /// microtubule-based movement /// cell cycle /// mitosis /// cell division	nucleotide bind. /// motor activity /// microtubule motor activity /// ATP bind.	nucleus /// endosome /// early endosome /// centrosome /// spindle /// microtubule /// microtubule associated complex
GSTO2	glutathione S-transferase omega 2	2.67E-04	2.3	227163_at	Hs.203634	metabolic process	glutathione transferase activity /// transferase activity	cytoplasm
HIST1H2BF	histone cluster 1, H2bf	2.90E-04	2.5	208490_x_at	Hs.182137	nucleosome assembly /// defense response to bacterium	DNA bind. /// protein bind.	nucleosome /// nucleus /// chromosome
RDM1	RAD52 motif 1	3.28E-04	2.0	239169_at	Hs.194411	---	nucleotide bind. /// nucleic acid bind. /// DNA bind. /// RNA bind.	nucleus /// cytoplasm
ANP32E	acidic (leucine-rich) nuclear phosphoprotein 32 family, member E	3.40E-04	2.7	208103_s_at	Hs.656466	---	protein bind. /// phosphatase inhibitor activity	nucleus /// cytoplasm /// cytoplasmic membrane-bounded vesicle
CSE1L	CSE1 chromosome segregation 1-like (yeast)	3.79E-04	2.2	201111_at	Hs.90073	protein import into nucleus, docking /// transport /// intracellular protein transport /// apoptosis /// cell prolif. /// protein transport	bind. /// protein bind. /// importin-alpha export receptor activity /// protein transporter activity	nucleus /// nuclear pore /// cytoplasm
UCA1	urothelial cancer associated 1	3.79E-04	8.7	227919_at	Hs.644234	---	---	---
UBE2T	ubiquitin-conjugating enzyme E2T (putative)	3.90E-04	2.6	223229_at	Hs.5199	modification-dependent protein catabolic process /// post-translational protein modification /// reg. of protein metabolic process	ubiquitin-protein ligase activity /// ligase activity /// small conjugating protein ligase activity	nucleus /// nucleolus /// cytoplasm
C3orf26	chromosome 3 open reading frame 26	4.02E-04	2.1	224523_s_at	Hs.655111	---	---	---
SELI	selenoprotein I	4.12E-04	2.8	1555274_a_at	Hs.189073	phospholipid biosynthetic process	magnesium ion bind. /// ethanolaminephosphotransferase activity /// selenium bind. /// transferase activity /// phosphotransferase activity, for other substituted phosphate groups /// manganese ion bind. /// metal ion bind.	membrane /// integral to membrane
GINS1	GINS complex subunit 1 (Psf1 homolog)	4.12E-04	2.9	206102_at	Hs.658464	inner cell mass cell prolif. /// DNA replication	protein bind.	nucleus /// cytoplasm
HSPA14	heat shock 70kDa protein 14	4.18E-04	2.1	227650_at	Hs.534169	response to stress	nucleotide bind. /// ATP bind.	---

Gene symbol	Gene title	Corr. *p*-value	FC	Probe set ID	UniGene ID	GO biological process term	GO molecular function term	GO cellular component term
CDC6	cell division cycle 6 homolog (S. cerevisiae)	4.24E-04	7.0	203968_s_at	Hs.405958	DNA replication checkpoint /// reg. of cyclin-dependent protein kinase activity /// DNA replication /// cell cycle /// mitosis /// traversing start control point of mitotic cell cycle /// neg. reg. of DNA replication /// neg. reg. of cell prolif. /// cell division	nucleotide bind. /// chromatin bind. /// protein bind. /// ATP bind. /// nucleoside-triphosphatase activity	nucleus /// nucleoplasm /// cytoplasm /// spindle /// cytosol
RFC4	replication factor C (activator 1) 4, 37kDa	4.28E-04	2.0	204023_at	Hs.714318	DNA replication /// DNA strand elongation during DNA replication /// DNA repair /// nucleotide-excision repair, DNA gap filling /// phosphoinositide-mediated signaling	nucleotide bind. /// DNA bind. /// DNA clamp loader activity /// protein bind. /// ATP bind. /// nucleoside-triphosphatase activity	nucleus /// nucleoplasm /// DNA replication factor C complex
KIF18A	kinesin family member 18A	4.28E-04	3.9	221258_s_at	Hs.301052	transport /// microtubule-based movement /// protein transport	nucleotide bind. /// motor activity /// microtubule motor activity /// ATP bind.	nucleus /// cytoplasm /// microtubule /// microtubule associated complex /// microtubule cytoskeleton /// cell projection
ATAD2	ATPase family, AAA domain containing 2	4.45E-04	2.2	235266_at	Hs.370834	transcription /// reg. of transcription, DNA-dependent /// reg. of transcription	nucleotide bind. /// DNA bind. /// transcription factor activity /// ATP bind. /// hydrolase activity /// nucleoside-triphosphatase activity /// ATPase activity, uncoupled	nucleus
RAD54B	RAD54 homolog B (S. cerevisiae)	4.47E-04	3.0	219494_at	Hs.30561	double-strand break repair via homologous recombination /// DNA repair /// mitotic recombination /// transcription /// reg. of transcription, DNA-dependent /// response to DNA damage stimulus /// reciprocal meiotic recombination /// response to ionizing radiation /// response to drug	nucleotide bind. /// nucleic acid bind. /// DNA bind. /// DNA helicase activity /// RNA helicase activity /// helicase activity /// protein bind. /// ATP bind. /// DNA translocase activity /// hydrolase activity	nucleus /// cytoplasm
TPX2	TPX2, microtubule-associated, homolog (Xenopus laevis)	4.49E-04	3.4	210052_s_at	Hs.719145	mitosis /// cell prolif.	protein bind. /// ATP bind. /// GTP bind.	spindle pole /// nucleus /// nucleolus /// spindle /// microtubule cytoskeleton
CDC6	cell division cycle 6 homolog (S. cerevisiae)	4.57E-04	8.7	203967_at	Hs.405958	DNA replication checkpoint /// reg. of cyclin-dependent protein kinase activity /// DNA replication /// cell cycle /// mitosis /// traversing start control point of mitotic cell cycle /// neg. reg. of DNA replication /// neg. reg. of cell prolif. /// cell division	nucleotide bind. /// chromatin bind. /// protein bind. /// ATP bind. /// nucleoside-triphosphatase activity	nucleus /// nucleoplasm /// cytoplasm /// spindle /// cytosol
HIST1H2AE	histone cluster 1, H2ae	4.92E-04	2.5	214469_at	Hs.121017	nucleosome assembly	DNA bind.	nucleosome /// nucleus /// chromosome
GMNN	geminin, DNA replication inhibitor	5.11E-04	2.4	218350_s_at	Hs.234896	cell cycle /// neg. reg. of DNA replication /// organ morphogenesis /// neg. reg. of cell cycle	protein bind.	nucleus /// nucleoplasm

Gene symbol	Gene title	Corr. *p*-value	FC	Probe set ID	UniGene ID	GO biological process term	GO molecular function term	GO cellular component term
KIF2C	kinesin family member 2C	5.28E-04	4.2	211519_s_at	Hs.69360	microtubule-based movement /// microtubule depolymerization /// mitosis /// cell prolif. /// establishment or maintenance of microtubule cytoskeleton polarity /// reg. of chromosome segregation	nucleotide bind. /// motor activity /// microtubule motor activity /// protein bind. /// ATP bind. /// centromeric DNA bind.	chromosome, centromeric region /// kinetochore /// condensed chromosome kinetochore /// nucleus /// cytoplasm /// cytoskeleton /// kinesin complex /// microtubule /// cytoplasmic microtubule /// microtubule cytoskeleton
---	---	5.42E-04	2.2	230294_at	---	---	---	---

Table 19: Annotation of 36 probe sets possessing putative prognostic value in human N+ intestinal-type gastric adenocarcinomas.

Welch-test was performed to indentify significantly differentially expressed genes between patients with and without recurrence of disease. Only genes with $p<0.001$ (very significant) and a fold difference (fold change) between group means of >1.5 were regarded as meaningful. Annotation information was obtained from NetAffx Analysis Center (09/2009). Genes are sorted with descending significance.

FC – fold change; recurr. – recurrence; ID – identifier; GO – Gene Ontology; reg. – reg.; activ. – activation; prolif. – proliferation; bind. – binding; pos. – positive; neg. – negative;

Gene symbol	Gene title	*p*-value	FC	Reg. in recurr.	Probe set ID	UniGene ID	GO biological process term	GO molecular function term	GO cellular component term
RANBP17	RAN bind. protein 17	8.77E-07	3.2	up	219661_at	Hs.410810	protein import into nucleus, docking /// protein import into nucleus /// transport /// intracellular protein transport /// protein transport /// mRNA transport /// intracellular protein transmembrane transport	bind. /// GTP bind. /// protein transporter activity	nucleus /// nuclear pore /// cytoplasm
HOXC10	homeobox C10	3.02E-05	13.9	up	218959_at	Hs.44276	skeletal system development /// transcription /// reg. of transcription, DNA-dependent /// multicellular organismal development /// pos. reg. of cell prolif. /// anterior/posterior pattern formation /// proximal/distal pattern format	DNA bind. /// transcription factor activity /// RNA polymerase II transcription factor activity /// transcription regulator activity /// sequence-specific DNA bind.	nucleus /// cyclin-dependent protein kinase activating kinase holoenzyme complex
---	---	3.04E-05	2.9	up	1559861_at	Hs.674409	---	---	---
---	---	4.75E-05	3.6	up	230406_at	Hs.662908	---	---	---
---	---	6.91E-05	2.1	up	233964_at	Hs.608498	---	---	---
VANGL2	vang-like 2 (van gogh, Drosophila)	7.30E-05	3.5	up	226029_at	Hs.99477	establishment of planar polarity /// neural tube closure /// heart looping /// multicellular organismal development /// sensory cilium assembly /// apical protein localization	protein bind.	membrane /// integral to membrane

Gene symbol	Gene title	*p*-value	FC	Reg. in recurr.	Probe set ID	UniGene ID	GO biological process term	GO molecular function term	GO cellular component term
WIPF2	WAS/WASL interacting protein family, member 2	1.62E-04	1.7	up	212051_at	Hs.421622	---	actin bind.	cytoplasm /// cytoskeleton
FOLR1	folate receptor 1 (adult)	1.99E-04	7.6	up	204437_s_at	Hs.73769	receptor-mediated endocytosis /// folic acid transport /// folic acid metabolic process	receptor activity /// folic acid bind.	extracellular region /// membrane fraction /// plasma membrane /// integral to plasma membrane /// membrane /// anchored to membrane
HIST1H4B	histone cluster 1, H4b	2.43E-04	2.4	up	232035_at	Hs.143080	establishment or maintenance of chromatin architecture /// nucleosome assembly /// phosphoinositide-mediated signaling	DNA bind. /// protein bind.	nucleosome /// nucleus /// chromosome
---	---	2.71E-04	2.7	up	230428_at	Hs.657711	---	---	---
TSLP	thymic stromal lymphopoietin	2.72E-04	2.5	up	235737_at	Hs.389874	---	cytokine activity	extracellular region /// extracellular space
RRAD	ras-related associated with diabetes	2.75E-04	2.6	up	204802_at	Hs.1027	small GTPase mediated signal transduction	nucleotide bind. /// GTPase activity /// protein bind. /// calmodulin bind. /// GTP bind.	---
DUOX1	dual oxidase 1	2.77E-04	3.5	up	219597_s_at	Hs.272813	response to oxidative stress /// cytokine-mediated signaling pathway /// cuticle development /// hormone biosynthetic process /// superoxide release /// hydrogen peroxide catabolic process /// hydrogen peroxide biosynthetic process /// response to cAMP /// oxidation reduction	peroxidase activity /// iron ion bind. /// calcium ion bind. /// electron carrier activity /// NAD(P)H oxidase activity /// oxidoreductase activity /// heme bind. /// FAD bind. /// NADP or NADPH bind.	plasma membrane /// membrane /// integral to membrane /// apical plasma membrane
MLLT3	myeloid/lymphoid or mixed-lineage leukemia (trithorax homolog, Drosophila); translocated to, 3	2.88E-04	2.3	down	1569652_at	Hs.591085	transcription /// reg. of transcription, DNA-dependent /// segment specification /// anterior/posterior pattern formation	---	nucleus /// cytoplasm
C8orf73	chromosome 8 open reading frame 73	3.08E-04	1.9	up	227672_at	Hs.531406	---	bind.	---
ZNF266	zinc finger protein 266	3.20E-04	1.6	down	214686_at	Hs.656185	transcription /// reg. of transcription, DNA-dependent	nucleic acid bind. /// DNA bind. /// zinc ion bind. /// metal ion bind.	intracellular /// nucleus

Gene symbol	Gene title	*p*-value	FC	Reg. in recurr.	Probe set ID	UniGene ID	GO biological process term	GO molecular function term	GO cellular component term
PERLD1	per1-like domain containing 1	3.37E-04	2.5	up	221811_at	Hs.462971	GPI anchor metabolic process /// GPI anchor biosynthetic process	hydrolase activity, acting on ester bonds	Golgi membrane /// endoplasmic reticulum /// endoplasmic reticulum membrane /// Golgi apparatus /// membrane /// integral to membrane /// intrinsic to endoplasmic reticulum membrane
MLF1	myeloid leukemia factor 1	3.57E-04	3.1	up	204783_at	Hs.85195	myeloid progenitor cell differentiation /// transcription /// cell cycle /// cell cycle arrest /// multicellular organismal development /// cell differentiation	DNA bind. /// protein bind. /// protein domain specific bind.	nucleus /// cytoplasm
ALDH3B1	aldehyde dehydrogenase 3 family, member B1	3.58E-04	3.0	up	205640_at	Hs.523841	cellular alcohol metabolic process /// cellular aldehyde metabolic process /// lipid metabolic process /// metabolic process /// oxidation reduction	3-chloroallyl aldehyde dehydrogenase activity /// aldehyde dehydrogenase [NAD(P)+] activity /// oxidoreductase activity	---
---	---	3.72E-04	3.3	up	234219_at	Hs.587370	---	---	---
S100A12	S100 calcium bind. protein A12	4.04E-04	3.5	up	205863_at	Hs.19413	xenobiotic metabolic process /// inflammatory response /// defense response to bacterium /// defense response to fungus	calcium ion bind. /// zinc ion bind. /// metal ion bind.	insoluble fraction /// nucleus /// nucleolus /// cytoplasm /// cytosol /// plasma membrane
DDIT3 /// NR1H3	DNA-damage-inducible transcript 3 /// nuclear receptor subfamily 1, group H, member 3	4.75E-04	1.5	up	209383_at	Hs.505777	response to amphetamine /// transcription /// reg. of transcription, DNA-dependent /// response to DNA damage stimulus /// response to oxidative stress /// ER overload response /// cell cycle /// cell cycle arrest /// aging /// response to nutrient /// cell death /// pos. reg. of specific transcription from RNA polymerase II promoter /// neg. reg. of foam cell differentiation /// endoplasmic reticulum unfolded protein response /// pos. reg. of cellular protein metabolic process /// neg. reg. of lipid transport /// pos. reg. of cholesterol transport /// neg. reg. of CREB transcription factor activity /// response to drug /// response to hydrogen peroxide /// mRNA transcription from RNA polymerase II promoter /// pos. reg. of apoptosis /// cell redox homeostasis /// pos. reg. of fatty acid biosynthetic process /// pos. reg. of transcription /// neg. reg. of pinocytosis /// embryonic organ development /// pos. reg. of lipoprotein lipase activity	nucleic acid bind. /// DNA bind. /// transcription factor activity /// steroid hormone receptor activity /// transcription coactivator activity /// transcription corepressor activity /// RNA bind. /// receptor activity /// ligand-dependent nuclear receptor activity /// ecdysteroid hormone receptor activity /// thyroid hormone receptor activity /// steroid bind. /// protein bind. /// zinc ion bind. /// sequence-specific DNA bind. /// metal ion bind. /// protein dimerization activity	intracellular /// nucleus /// nucleolus /// cytoplasm

Gene symbol	Gene title	*p*-value	FC	Reg. in recurr.	Probe set ID	UniGene ID	GO biological process term	GO molecular function term	GO cellular component term
FBLIM1	filamin bind. LIM protein 1	4.85E-04	1.8	up	1555483_x_at	Hs.530101	cell adhesion /// reg. of cell shape	protein bind. /// zinc ion bind. /// metal ion bind.	cytoplasm /// cytoskeleton /// cell junction
RASAL1	RAS protein activator like 1 (GAP1 like)	5.69E-04	2.5	up	219752_at	Hs.528693	signal transduction /// intracellular signaling cascade /// reg. of small GTPase mediated signal transduction	GTPase activator activity /// Ras GTPase activator activity /// phosphor-lipid bind. /// zinc ion bind. /// metal ion bind.	intracellular
---	---	5.70E-04	2.4	up	242799_at	Hs.445931	---	---	---
SOX7	SRY (sex determining region Y)-box 7	5.97E-04	1.8	up	224013_s_at	Hs.709543	transcription /// reg. of transcription, DNA-dependent /// reg. of transcription from RNA polymerase II promoter	DNA bind. /// transcription factor activity	nucleus
RRM2	ribonucleotide reductase M2 polypeptide	6.41E-04	1.8	down	209773_s_at	Hs.226390	DNA replication /// deoxyribonucleo-side diphosphate metabolic process /// deoxyribonucleotide metabolic process /// deoxyribonucleotide biosynthetic process /// protein oligomerization /// oxidation reduction	ribonucleoside-diphosphate reductase activity /// iron ion bind. /// protein bind. /// oxidoreductase activity /// metal ion bind. /// transition metal ion bind.	cytoplasm /// cytosol
RRM2	ribonucleotide reductase M2 polypeptide	6.54E-04	1.8	down	201890_at	Hs.226390	DNA replication /// deoxyribonucleo-side diphosphate metabolic process /// deoxyribonucleotide metabolic process /// deoxyribonucleotide biosynthetic process /// protein oligomerization /// oxidation reduction	ribonucleoside-diphosphate reductase activity /// iron ion bind. /// protein bind. /// oxidoreductase activity /// metal ion bind. /// transition metal ion bind.	cytoplasm /// cytosol
LOC100133772 /// SLC16A5	similar to MCT /// solute carrier family 16, member 5 (monocarbox-ylic acid trans-porter 6)	7.08E-04	2.2	up	206599_at	Hs.592095	transport /// organic anion transport /// monocarboxylic acid transport	monocarboxylic acid transmembrane transporter activity /// symporter activ-ity /// secondary active monocarboxy-late transmembrane transporter activity	membrane fraction /// plasma mem-brane /// integral to plasma membrane /// membrane /// integral to membrane
AKR1C1	aldo-keto re-ductase family 1, member C1 (dihydrodiol dehydrogenase 1; 20-alpha (3-alpha)-hydroxysteroid dehydrogenase)	7.29E-04	3.2	up	217626_at	Hs.460260	lipid metabolic process /// pros-taglandin metabolic process /// xenobi-otic metabolic process /// digestion /// steroid metabolic process /// bile acid metabolic process /// bile acid and bile salt transport /// intestinal cholesterol absorption /// cholesterol homeostasis /// protein homooligomerization /// oxidation reduction	aldo-keto reductase activity /// oxi-doreductase activity /// carboxylic acid bind. /// bile acid bind. /// 20-alpha-hydroxysteroid dehydrogenase activity /// 3-alpha-hydroxysteroid dehydro-genase (A-specific) activity /// 3-alpha-hydroxysteroid dehydrogenase (B-specific) activity /// trans-1,2-dihydrobenzene-1,2-diol dehydro-genase activity	cytoplasm /// cytosol

Gene symbol	Gene title	*p*-value	FC	Reg. in recurr.	Probe set ID	UniGene ID	GO biological process term	GO molecular function term	GO cellular component term
GALNT5	UDP-N-acetyl-alpha-D-galactosamine:polypeptide N-acetylgalactosaminyltransferase 5 (GalNAc-T5)	7.52E-04	1.8	up	240390_at	Hs.269027	glycosaminoglycan biosynthetic process	polypeptide N-acetylgalactosaminyltransferase activity /// calcium ion bind. /// sugar bind. /// transferase activity /// transferase activity, transferring glycosyl groups /// manganese ion bind.	Golgi membrane /// Golgi apparatus /// membrane /// integral to membrane
ALS2CL	ALS2 C-terminal like	8.73E-04	1.6	up	229887_at	Hs.517937	endosome organization /// protein localization	GTPase activator activity /// protein bind. /// Rab GTPase bind. /// identical protein bind.	cytoplasm /// cytoplasmic membrane-bounded vesicle
---	---	8.79E-04	2.6	up	1563621_at	Hs.681809	---	---	---
FAM53A	family with sequence similarity 53, member A	8.98E-04	1.8	up	1569139_s_at	Hs.143314	---	---	nucleus
LOC283070	hypothetical LOC283070	9.35E-04	2.2	down	226382_at	Hs.600547	---	---	---
---	---	9.59E-04	1.6	up	1556172_at	Hs.132305	---	---	---

Table 20: Annotation of 18 probe sets possessing putative prognostic value in human N+ diffuse-type gastric adenocarcinomas.

Welch-test was performed to indentify significantly differentially expressed genes between patients with and without recurrence of disease. Only genes with $p<0.001$ (very significant) and a fold difference (fold change) between group means of >1.5 were regarded as meaningful. Annotation information was obtained from NetAffx Analysis Center (09/2009). Genes are sorted with descending significance.

FC – fold change; recurr. – recurrence; ID – identifier; GO – Gene Ontology; reg. – reg.; activ. – activation; prolif. – proliferation; bind. – binding; pos. – positive; neg. – negative

Gene symbol	Gene title	*p*-value	FC	Reg. in recurr.	Probe set ID	UniGene ID	GO biological process term	GO molecular function term	GO cellular component term
TXNDC11	thioredoxin domain containing 11	7.68E-05	1.5	down	223325_at	Hs.313847	cell redox homeostasis	---	endoplasmic reticulum /// endoplasmic reticulum membrane /// membrane /// integral to membrane

Gene symbol	Gene title	*p*-value	FC	Reg. in recurr.	Probe set ID	UniGene ID	GO biological process term	GO molecular function term	GO cellular component term
HOPX	HOP homeobox	7.69E-05	3.2	up	211597_s_at	Hs.654864	neg. reg. of transcription from RNA polymerase II promoter /// trophectodermal cell differentiation /// trophectodermal cell differentiation /// transcription /// reg. of transcription, DNA-dependent /// multicellular organismal developmen	DNA bind. /// transcription factor activity /// protein bind. /// transcription repressor activity /// sequence-specific DNA bind.	nucleus
KIAA0746	KIAA0746 protein	9.72E-05	1.6	down	212314_at	Hs.479384	---	bind.	membrane /// integral to membrane
EPHA4	Ephrin receptor A4	1.13E-04	2.3	up	227449_at	Hs.371218	protein amino acid phosphorylation /// signal transduction /// transmembrane receptor protein tyrosine kinase signaling pathway /// axon guidance /// adult walking behavior	nucleotide bind. /// protein kinase activity /// protein tyrosine kinase activity /// transmembrane receptor protein tyrosine kinase activity /// receptor activity /// ephrin receptor activity /// protein bind. /// ATP bind. /// kinase activity /// transferase activity	integral to plasma membrane /// membrane /// integral to membrane
MYADM	myeloid-associated differentiation marker	1.94E-04	1.7	up	225673_at	Hs.380906	---	---	membrane /// integral to membrane
BTN3A2 /// BTN3A3	butyrophilin, subfamily 3, member A2 /// butyrophilin, subfamily 3, member A3	2.58E-04	1.6	down	204820_s_at	Hs.376046	---	---	membrane /// integral to membrane
KIAA0746	KIAA0746 protein	3.36E-04	1.8	down	212311_at	Hs.479384	---	bind.	membrane /// integral to membrane
RND3	Rho family GTPase 3	3.40E-04	1.6	up	212724_at	Hs.6838	cell adhesion /// small GTPase mediated signal transduction /// actin cytoskeleton organization	nucleotide bind. /// GTPase activity /// GTP bind.	Golgi membrane /// intracellular /// Golgi apparatus /// membrane
BTN3A2	butyrophilin, subfamily 3, member A2	3.67E-04	1.6	down	209846_s_at	Hs.376046	---	---	membrane /// integral to membrane
---	---	3.93E-04	2	up	232573_at	Hs.667730	---	---	---
ST6GAL1	ST6 beta-galactosamide alpha-2,6-sialyltranferase 1	4.97E-04	1.9	down	201998_at	Hs.207459	protein modification process /// protein amino acid glycosylation /// humoral immune response /// oligosaccharide metabolic process	beta-galactoside alpha-2,6-sialyltransferase activity /// sialyltransferase activity /// transferase activity /// transferase activity, transferring glycosyl groups	extracellular region /// Golgi apparatus /// membrane /// integral to membrane /// integral to Golgi membrane

Gene symbol	Gene title	*p*-value	FC	Reg. in recurr.	Probe set ID	UniGene ID	GO biological process term	GO molecular function term	GO cellular component term
SGCD	sarcoglycan, delta (35kDa dystrophin-associated glycoprotein)	7.03E-04	1.6	up	210330_at	Hs.387207	cytoskeleton organization /// muscle organ development	protein bind.	cytoplasm /// cytoskeleton /// plasma membrane /// sarcoglycan complex /// membrane /// integral to membrane
PDLIM4	PDZ and LIM domain 4	7.18E-04	2.1	up	211564_s_at	Hs.424312	---	protein bind. /// zinc ion bind. /// metal ion bind.	---
RBPMS /// SDHALP2	RNA bind. protein with multiple splicing /// succinate dehydrogenase complex, subunit A, flavoprotein pseudogene 2	7.19E-04	2.1	up	1557223_at	Hs.334587	RNA processing	nucleotide bind. /// nucleic acid bind. /// RNA bind. /// protein bind. /// electron carrier activity /// oxidoreductase activity	---
BTN3A2	butyrophilin, subfamily 3, member A2	7.29E-04	1.7	down	212613_at	Hs.376046	---	---	membrane /// integral to membrane
CREB5	CAMP responsive element bind. protein 5	8.38E-04	3.2	up	232555_at	Hs.437075	transcription /// reg. of transcription, DNA-dependent /// transcription from RNA polymerase II promoter /// pos. reg. of transcription, DNA-dependent	DNA bind. /// transcription factor activity /// protein bind. /// zinc ion bind. /// sequence-specific DNA bind. /// metal ion bind. /// protein dimerization activity	intracellular /// nucleus
NCRNA 00107	non-protein coding RNA 107	8.82E-04	1.7	down	232463_at	Hs.575741	---	---	---
GAP43	growth associated protein 43	8.85E-04	3.8	up	204471_at	Hs.134974	activation of protein kinase C activity by G-protein coupled receptor protein signaling pathway /// multicellular organismal development /// nervous system development /// axon guidance /// response to wounding /// glial cell differentiation /// axon choice point recognition	protein bind. /// calmodulin bind.	plasma membrane /// membrane /// cell junction /// axon /// cell projection /// synapse

List of abbreviations, symbols and dimensions

ACTA2	Actin, alpha 2, smooth muscle, aorta (α-smooth muscle actin)
ACTB	β-actin gene/transcript
AEC	3-amino-9-ethylcarbazole
AP	alkaline phosphatase
ATP	adenosine triphosphate
AUT	Austria
BAMBI	BMP and activin membrane-bound inhibitor homolog
BAMBI	BMP and activin membrane-bound inhibitor homolog gene/transcript
BMP	bone morphogenetic protein
bp	base pairs
°C	degrees celsius (centigrade)
c	concentration
CAN	Canada
Cat.	catalogue
cDNA	complementary DNA
DEN	Denmark
C-terminus	carboxy-terminus
Da	Dalton
DAPI	4',6-diamidino-2-phenylindole-dihydrochloride
DEPC	diethylpyrocarbonate
ddH_2O	double-distilled water
Dig	digoxigenin
DMEM	Dulbecco's Modified Eagle Medium

DMSO	dimethylsulfoxid
DNA	deoxyribonucleic acid
DTT	dithiothreitol
EDTA	ethylenediaminetetraacetic acid
e.g.	*exempli gratia* (for example)
etc.	*et cetera* (and so forth)
EPHA4	Ephrin receptor A4
EPHA4	Ephrin receptor A4 gene/transcript
EU	European Union
fab	fragment antigen binding
FBS	fetal bovine serum
FDR	False Discovery Rate
Fig.	figure
FOLR1	folate receptor 1
FOLR1	folate receptor 1 gene/transcript
FWER	Family Wise Error Rate
g	gram
g	strength of the earth's gravitational field (g=9.81 m/s^2)
GAP43	growth associated protein 43 gene/transcript
GAPDH	glyceraldehyde-3-phosphate dehydrogenase gene/transcript
GC	guanine cytosine
GER	Germany
GmbH	"*Gesellschaft mit beschränkter Haftung*"
HE	hematoxylin and eosin
HG	human genome
H_2O	water

H_2O_2	hydrogen peroxide
HOXC10	homeobox C10 gene/transcript
IARC	International Agency for Research on Cancer
Ig	immunoglobulin
Inc.	Incorporated
IVT	*in vitro* transcription
JPN	Japan
k	kilo (×1,000)
KRT	keratin (cytokeratin)
l	liter
M	molar
m	meter
m (prefix)	10^{-3} (milli)
max.	maximal/maximum
MEM	minimum essential medium
min	minute/s
mol	mol
mRNA	messenger ribonucleic acid
MW	molecular weight
N/A	not available
n (prefix)…	10^{-9} (nano)
n	number
NEAA	non essential amino acids
N-terminus	amino-terminus
dNTP	deoxyribonucleotide
NTMT	NaCl-Tris-HCl-magnesium chloride-Tween

p.a.	pro analysi
PBS	phosphate buffered saline
PCR	polymerase chain reaction
RANBP17	RAN binding protein 17
RANBP17	RAN binding protein 17 gene/transcript
RIN	RNA (ribonucleic acid) integrity number
RNA	ribonucleic acid
rpm	rotations per minute
RPMI	Roswell Park Memorial Institute
RRAD	ras-related associated with diabetes
RRAD	ras-related associated with diabetes gene/transcript
RZPD	*"Ressourcenzentrum Primaerdatenbank"/"Deutsches Ressourcenzentrum fuer Genomforschung"*
s	second/s
S	Svedberg
SDS	sodium dodecyl sulfate
*α*SMA	*α*-smooth muscle actin
SMAD	human homolog of *caenorhabditis elegans* "SMA" and *drosophila* "mothers against decapentaplegic" (MAD)
SSC	saline sodium citrate
Tab.	table
TBS	Tris-buffered saline
TCF-4	T-cell factor 4 (official gene symbol: TCF7L2)
TCF7L2	transcription factor 7-like 2 (T-cell-specific, HMG-box)
TE	Tris-HCl-EDTA
TES	Tris-HCl-EDTA-SDS
TGF	transforming growth factor

THBS4	thrombospondin 4
THBS4	thrombospondin 4 gene/transcript
Tris	tris(hydroxymethyl) aminomethane
U	units
UICC	*Union Internationale Contre Le Cancer*
USA	United States of America
µ (prefix)	10^{-6} (micro)
VIM	vimentin
RT	reverse transcriptase
x	times (multiplied by)
×	n-fold
#	number
~/≈	approximate/ly

Danksagung

Mein spezieller Dank gebührt Dr. habil. Wolfgang Kemmner (Forschungsgruppe *Surgical Oncology*, Experimental and Clinical Research Center, Berlin) für die Überlassung dieses höchst spannenden Forschungsthemas, dessen Bearbeitung mich fortlaufend vor neue Herausforderungen gestellt hat. Ich möchte mich bei ihm für das mir entgegengebrachte Vertrauen und die mir eingeräumten Freiräume, eigene wissenschaftlichen Ideen und Hypothesen zu entwickeln und zu bearbeiten, bedanken. Dies ermöglichte es mir, selbständig und eigenverantwortlich zu arbeiten, was mich mit großer wissenschaftlicher und persönlicher Erfüllung bereicherte. Er stand mir als Ansprechpartner bei Problemen und für kritische wissenschaftliche Diskussionen immer zur Verfügung.

Mein Dank gilt Prof. Dr. Hanspeter Herzel (Institut für Theoretische Biologie; Humboldt-Universität zu Berlin), der es mir ermöglichte, meine Dissertation am Institut für Biologie der Humboldt-Universität zu Berlin durchzuführen. Er stand mir mit Rat und Tat bei Problematiken der bioinformatischen Microarray-Auswertung zur Seite.

I owe a debt of gratitude to Masakazu Yashiro (M.D., Ph.D.) from the Department of Surgical Oncology, Osaka City University Graduate School of Medicine for providing the human fibroblast cultures, and for exchange of ideas and fruitful discussions.

Weiterer Dank gilt Dr. Qing Wang (Forschungsgruppe *Surgical Oncology*, Experimental and Clinical Research Center, Berlin) für ihre immerwährende Unterstützung auf wissenschaftlicher und emotionaler Ebene, Sabine Grigull (Forschungsgruppe *Surgical Oncology*, Experimental and Clinical Research Center, Berlin) für fortlaufende Hilfestellungen in der Zellkultur und Dr. Christian Klein (Forschungsgruppe *Angiogenesis and Cardiovascular Pathology*, Max-Delbrück-Centrum für Molekulare Medizin, Berlin) für kontinuierlichen kritischen Gedankenaustausch.

Das erfolgreiche Gelingen der quantitativen *real-time* PCRs sowie anderer molekularbiologischer Methoden ist dem fortwährenden Gedankenaustausch mit Dr. Daniel Militz (Forschungsgruppe *Molecular Cardiovascular Research*, Max-Delbrück-Centrum für Molekulare Medizin, Berlin) maßgeblich zuzuschreiben. Hierfür gilt ihm mein ganz spezieller und persönlicher Dank.

Weiteren Dank möchte ich Dr. Thomas Jöns und Carola Meier (Zentrum für Anatomie, Charité-Universitätsmedizin, Berlin) für Ihre Hilfe bei immunhistochemischen Methoden aussprechen.

PD Dr. Michael Vieth (Institut für Pathologie, Klinikum Bayreuth GmbH, Bayreuth) möchte ich für das kritische Lesen der Einleitung dieser Doktorarbeit und für die Unterstützung bei pathologischen Fragestellungen danken.

Dr. Johannes Fritzmann (Forschungsgruppe *Signal Transduction, Invasion and Metastasis of Epithelial Cells*, Max-Delbrück-Centrum für Molekulare Medizin, Berlin) gilt mein Dank für die Bereitstellung der *BAMBI*-Sonde sowie für ausgiebigen Gedankenaustausch bezüglich *Wnt/β-catenin* - und *TGF-β signaling* und klinisch-onkologischen Fragestellungen.

PD Dr. Stephan Gretschel (Klinik für Allgemein-, Visceral-, Gefäß- und Thoraxchirurgie, Charité-Universitätsmedizin, Berlin) möchte ich für die fortwährende Hilfe bei klinisch-onkologischen Problemen danken.

Die Unterstützung durch Dr. Frank Kleinjung (MicroDiscovery GmbH, Berlin), der mir mit unglaublicher Geduld bei der Lösung statistischer und bioinformatischer Probleme zur Seite stand, möchte ich in diesem Rahmen ebenfalls unbedingt anerkennen.

Nicht zuletzt möchte ich meine Eltern, Dr. habil. Andrea Förster und Dr. habil. Hans-Jürgen Förster, erwähnen, ohne deren fortwährende Hilfe und Unterstützung – wissenschaftlicher und privater Natur – meine erfolgreiche Promotion wohl unter einem anderen Stern gestanden hätte.

Lebenslauf

Der Lebenslauf wird aus datenschutzrechtlichen Gründen in der elektronischen Version dieser Arbeit nicht veröffentlicht.

Selbständigkeitserklärung

Hiermit erkläre ich, dass ich die vorliegende Arbeit mit dem Titel "*Gene expression profiling of human lymph node-positive gastric adenocarcinomas: Towards personalized prognosis and therapy*" selbständig und ausschließlich unter Verwendung der angegebenen Hilfsmittel und Literatur angefertigt habe.

Außerdem erkläre ich, dass ich mich nicht anderweitig um einen entsprechenden Doktorgrad beworben habe und einen solchen auch nicht besitze.

Die Promotionsordnung der Mathematisch-Naturwissenschaftlichen Fakultät I habe ich zur Kenntnis genommen und akzeptiert.

Susann Förster

Berlin, 04.03.2010

Publikationen

Wissenschaftliche Zeitschriften (*ISI Journals*)

Behrens, M., **Foerster, S.**, Staehler, F., Raguse, J.-D. and Meyerhof, W. (2007): Gustatory Expression Pattern of the Human TAS2R Bitter Receptor Gene Family Reveals a Heterogenous Population of Bitter Responsive Taste Receptor Cells. ─ *The Journal of Neuroscience*, vol. 27, no. 46, p. 12630–12640.

http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=18003842

Konferenzbeiträge

Förster, S., Kemmner W. (2009): Thrombospondin – a tumor stroma constituent of diffuse gastric cancer potentially important for progression and invasion. In: *20th European Students' Conference*, Berlin, Germany, October 4–7, p. 44

Brockhoff, A., Behrens, M., **Förster, S.**, Reichling, C., Bufe, B., Kuhn, C., Raguse, J.-D. and Meyerhof, W. (2006): Molecular Biology of Human Bitter Taste Receptors. In: *The 40th Annual Meeting of the Japanese Association for the Study of Taste and Smell (JATS)*, Fukuoka, Japan; July 11–13, Symposium 1, p. 4.

Förster, S., Parlitz, S., Kuhn, C., Behrens, M., Winnig, M., Bufe, B. and Meyerhof, W. (2005): Expression profiles of human bitter taste receptor genes. In: *25th Blankenese Conference "Signalling in Sensory Systems"*, Hamburg Blankenese, Germany, May 21–25, p. 52.

Diplomarbeit

Förster, S. (2006): Analyse der Expressionsmuster humaner Bitterrezeptoren im gustatorischen System – Universitäts-Bibliothek, Universität Potsdam, Potsdam, Deutschland

Susann Förster

Berlin, 04.03.2010